Draft Technology Assessment

Number xx

Continuous Positive Airway Pressure Treatment for Obstructive Sleep Apnea

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. To be included in the final report.

Prepared by:
To be included in the final report.

Investigators:
To be included in the final report.

AHRQ Publication No. xx-EHCxxx
<Month Year

This report is based on research conducted by the <EPC> Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. XXX-20XX-XXXXX). 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.

Suggested citation: To be included in the final report.

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: <EPC> Evidence-based Practice Center (Contract Number: XXX-20XX-XXXXX).

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

Christine Chang, M.D., M.P.H.
Acting 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

Acknowledgments

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

To be included in the final report.

Peer Reviewers

To be included in the final report.

Continuous Positive Airway Pressure Treatment for Obstructive Sleep Apnea

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, are typically assessed using the apnea-hypopnea index (AHI). However, no standard definition of this measure exists, and whether AHI (and associated measures) are valid surrogate measure of clinical outcomes is unknown. 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 clinical 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, Cochrane databases, CINAHL, and ClinicalTrials.gov from January 2010 through November 18, 2019; 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 randomized controlled trials (RCT) and adjusted nonrandomized comparative studies (NRCS) of CPAP and other comparative studies that reported both changes in potential intermediate or surrogate measures (e.g., AHI) and effects on clinical outcomes. All studies had to report effects on long-term (≥6 or 12 months) clinical outcomes in adults with OSA.

Results. The 47 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 the 25 studies that compared CPAP and no CPAP (n=23) or sham CPAP (n=2), 12 were RCTs and 13 NRCSs; 14 were analyzed as intention-to-treat (ITT) and 11 compared CPAP users to nonusers (either never-users or noncompliant users). All outcomes of interest were addressed by RCTs; the NRCSs mostly addressed composite cardiovascular (CV) outcomes and death.

RCTs provide low strength of evidence (SoE) that CPAP does not affect the risk of all-cause mortality (summary effect size [ES] 0.87, 95% confidence interval [CI] 0.58 to 1.29), stroke (summary ES 0.96, 95% CI 0.59 to 1.29), myocardial infarction (summary ES 1.06, 95% CI 0.72 to 1.56), or composite CV outcomes (ES range 0.42 to 1.10 across studies, all statistically nonsignificant). Regarding all-cause mortality, NRCSs were consistent with RCTs in direction of association. When NRCSs were combined with the RCTs there was low SoE that CPAP reduces risk of mortality (ES 0.66, 95% CI 0.60 to 0.73); although this conclusion may be most applicable to older adults and longer-term followup. RCTs provided insufficient evidence regarding risk of CV death, but combined with a NRCS, there is low SoE of no effect of CPAP (ES 0.97, 95% CI 0.62 to 1.53). NRCSs did not alter conclusions regarding other CV-related

outcomes. Insufficient evidence exists regarding effect of CPAP on the risk of transient ischemic attack, angina, coronary artery revascularization, congestive heart failure, and atrial fibrillation.

Regarding other assessed outcomes, CPAP does not affect the risk of driving accidents or the risk of incident diabetes (both low SoE). CPAP does not result in clinically significant changes in depression or anxiety scores, executive cognitive function measures, or nonspecific quality of life measures (all low SoE). There is insufficient evidence regarding the effect of CPAP on incident hypertension, functional status measures, male or female sexual function, or days of work missed.

Insufficient evidence exists regarding possible differences in the effect of CPAP on various outcomes based on patient characteristics (such as disease severity or comorbidities), different diagnostic criteria, or whether studies were analyzed as ITT or "as-treated". No study reported within-study correlations among outcomes (e.g., the association between the effects of CPAP on AHI and on all-cause mortality).

Eligible studies provided insufficient evidence about adverse events due to CPAP use. Adverse events reported in the Food and Drug Administration database mostly related to inadequate humidification, user errors, or device malfunction. No deaths were attributed to CPAP use.

No studies have evaluated the validity of intermediate or surrogate measures (such as change in AHI) as predictors of long-term clinical outcomes, including surrogacy or mediation analyses. Studies did not compare the concordance of different polysomnography and symptom measures with clinical outcomes. Across the 15 studies that reported both changes in intermediate or surrogate measures and effects on clinical outcomes, data were too sparse to allow adequate cross-study evaluation of concordance between any specific measure and clinical outcome.

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 AHI as a surrogate or intermediate outcome for long-term clinical 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 clinical outcomes.

The published evidence mostly does not support 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 clinically important outcomes. When NRCSs are combined with the RCTs there is the suggestion that CPAP reduces the risks of all-cause mortality (low SoE); other conclusions are not changed. The low SoE for these outcomes suggests that we have limited confidence that the summary estimates are close to the true effect.

Studies did not adequately address whether effects of CPAP vary 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 studies are needed before we have a clear understanding of 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 or whether of AHI (and/or other breathing measures) are valid intermediate or surrogate measures of clinical outcomes.

Contents

Tables

Figures

Appendixes

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 Functional Outcomes Sleep Questionnaire (FOSQ) is an OSA-specific functional status tool that is commonly used, particularly in research settings, to assess health status of people diagnosed with OSA.
  - 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 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, 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, treatment has aimed to prevent these outcomes.
- **Definitions of breathing measures used across studies**
  - Studies are highly inconsistent in the criteria used to define polysomnography 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.
  - Most studies (60%) did not fully and explicitly report the definitions of polysomnography measures used.
  - 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 clinical 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.
- **Clinical effect of CPAP versus no CPAP**
  - 25 studies (12 RCTs and 13 NRCSs) compared CPAP with no CPAP, of which 14 reported intention-to-treat (ITT) analyses and 11 as-treated analyses. Two RCTs used sham CPAP as the comparator.
  - Few studies were explicitly powered for long-term clinical outcomes: 2 RCTs were designed to be powered for composite CV outcomes (SAVE, RICCADSA); however, RICCADSA was ultimately underpowered. One RCT (APPLES) was powered for neurocognitive outcomes. No study was explicitly powered for any other outcome. 14 studies (56%) reported intention-to-treat analyses; 11 (44%) conducted as-treated analyses of compliant CPAP users. The as-treated analyses may not have been able to fully account for possible biases related to self-selection regarding CPAP use and compliance.
  - 3 RCTs provide an imprecise estimate of whether CPAP alters the risk of all-cause mortality (thus, a conclusion of low SoE of no effect). Additional evidence from 5 adjusted NRCSs were consistent with the RCTs in direction of association, but were statistically significant. Compared with the RCTs, the NRCSs had higher death rates related to older age (in 2 NRCSs) and longer-term followup (up to 11 years). Combining the RCTs and NRCSs provided low SoE that CPAP reduces the risk of all-cause mortality; this conclusion may be most applicable to older adults and longer-term followup. Omitting the one very large NRCS did not alter the conclusion. The ITT analyzed studies (in which compliance ranged from 38% to 60%) found similar associations as the as-treated studies (of compliant users). No definitive explanation was provided for the lack of difference in effect between analyses with and without noncompliant CPAP users.
  - 3 RCTs provide insufficient evidence regarding the effect of CPAP on risk of cardiovascular (CV) mortality, but together with 1 NRCS, there was low SoE that CPAP does not affect the risk of CV death.
  - 4 RCTs provide low SoE that CPAP does not affect risk of stroke or of acute myocardial infarction (MI).
  - 6 RCTs provide mostly imprecise estimates of whether CPAP alters the risk of various composite CV outcomes (thus, a conclusion of low SoE of no effect). Additional evidence from 3 adjusted NRCSs did not change this conclusion. Across studies, the as-treated analyses of CPAP users did not find significantly different associations than the ITT analyses of assignment to CPAP use (in which compliance ranged from about 38% to 64%).
  - 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 atrial fibrillation (AFib). For each outcome, there were few studies and/or effect estimates were highly imprecise.
  - 2 RCTs provide low SoE that CPAP does not affect the likelihood of driving accidents after 1 or about 3 to 4 years of use.
  - 2 RCTs provide an imprecise estimate of whether CPAP alters the risk of incident type 2 diabetes (thus, a conclusion of low SoE of no effect). An additional adjusted NRCS, though statistically significant, does not alter the conclusion.
  - There is low SoE that CPAP does not result in clinically meaningful changes in depression symptoms (based on 4 RCTs), anxiety symptoms (3 RCTs), executive cognitive function (3 RCTs and 1 NRCS), or quality of life (QoL, 8 RCTs and 1 NRCS); although the small differences were statistically significant.
  - 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.
  - The studies do not provide evidence that CPAP use is more (or less) effective in any specific subgroup. For all clinical 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 polysomnography measures, or whether ITT or as-treated analyses were reported. ITT studies were not consistently different in findings than as-treated analyses of compliant CPAP users. Within the one RCT that evaluated the association of compliance and outcomes, and across studies, there is no evidence of correlation between rate of compliance and CV outcomes.
  - No study reported within-study correlations among outcomes (e.g., effect on AHI and effect on all-cause mortality).
  - Based on the evaluated long-term, comparative studies, there is insufficient evidence regarding the risk of adverse events related to CPAP use.
- **Clinical effect of CPAP versus other active treatments**
  - There is low SoE that changes in depression and anxiety symptoms do not significantly differ between patients receiving either CPAP or MAD. The comparative effects on QoL, functional status, and sexual function are insufficient; other long-term clinical outcomes have not been reported.
  - There is moderate SoE that changes in functional status scores do not significantly differ between patients prescribed either fixed CPAP or autoCPAP. Other long-term clinical outcomes have not been reported.
- **Adverse events related to CPAP**
  - Studies that reported comparative, long-term clinical outcomes provide insufficient evidence regarding adverse events.
  - Adverse events reported in the Food and Drug Administration 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**
  - Studies rarely evaluated potential differences in effect across subgroups of participants. Both within and across studies, there is insufficient evidence to determine potential heterogeneity of any putative treatment effect.
- **Validity of intermediate or surrogate measures and clinical outcomes**
  - No study directly evaluated the validity of changes in breathing or sleepiness measures an intermediate or surrogate measures for long-term clinical outcomes. No study explicitly evaluated surrogacy or mediation analyses of the measures.
  - Studies did not evaluate how different polysomnography and symptom measures compared in their concordance with clinical outcomes.
  - Too few studies reported on any given pair of breathing or sleepiness measures and long-term clinical outcomes to allow adequate cross-study evaluation of concordance.
- **Research gaps**
  - Additional adequately powered, high-quality, long-term RCTs are needed to evaluate the effect of CPAP on clinical outcomes. Ideally, RCTs should be powered to evaluate potential differential effects of CPAP based on such factors as baseline AHI, 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). Some of these gaps could be addressed by re-analyses of existing RCTs; although, as with all secondary analyses, the analyses might be considered to be only hypothesis-generating. Individual participant-level meta-analysis of existing RCTs could also address these gaps.
  - Studies are needed to assess the validity of AHI, other breathing measures, and sleepiness scores as intermediate or surrogate measures for long-term clinical outcomes. Ideally, studies should assess and compare multiple putative intermediate or surrogate measures. Existing studies or databases may have sufficient data to allow such exploratory analyses.

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. Obstructive sleep apnea (OSA) is the most common type of sleep apnea. In OSA, the volume of airflow is diminished despite appropriate ventilator effort. It differs from central sleep apnea, in which there is a reduced drive to breathe, reduced ventilator 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 apnea-hypopnea index (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 terms of which patients are included and how treatments are provided. This may hinder the interpretation of the studies.

The most common first-line therapy for OSA is the use of continuous positive airway pressure (CPAP) devices during sleep. The CPAP machine 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).

This systematic review evaluates: (1) the variability across research studies in definitions of breathing measures (e.g., apneas, hypopneas) and criteria to diagnose OSA; (2) the effectiveness, comparative effectiveness, and harms of CPAP use on long-term clinically significant outcomes; and (3) the validity of AHI and similar measures as a surrogate or intermediate measure for clinically significant outcomes.

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 November 18, 2019 (literature search update pending).

Results

Definitions of breathing measures: Across 47 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 41 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 (26/30) 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.

CPAP versus no CPAP: 12 randomized controlled trials (RCTs; total N = 6019) and 13 nonrandomized comparative studies (NRCSs) with adjustment for potential confounding (N = 32,062; 25,389 in one study that reported only on all-cause mortality, 6673 in other NRCSs) reported long-term clinical 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; 3 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 25 studies, 14 studies (56%) reported intention-to-treat analyses; 11 (44%) conducted as-treated analyses of compliant CPAP users. The as-treated analyses may not have been able to fully account for possible biases related to self-selection regarding CPAP use and compliance.

Three RCTs yielded an imprecise estimate of the effect of CPAP on all-cause mortality (effect size 0.87, 95% CI 0.58 to 1.29); 5 NRCSs reporting on all-cause mortality had

associations in the same direction as the RCTs, but with stronger effect sizes that were more likely to be 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). Based on power considerations, the higher event rates in the NRCSs increased the likelihood of finding statistically significant effect sizes. Combining the RCTs and NRCSs yielded a statistically significant effect size (0.66, 95% CI 0.60 to 0.73), favoring CPAP. Compliance in the ITT analyses ranged from an estimate of about 38 to 60 percent. One RCT found a stronger, but still statistically nonsignificant effect in their compliant CPAP user analyses compared with the ITT analyses. This RCT also found no association between degree of CPAP compliance and CV outcomes. Across studies, the summary effect sizes of ITT and as-treated analyses were nearly identical.

Three RCTs yielded a highly imprecise estimate of the effect of CPAP on CV mortality (0.99, 95% CI 0.43 to 2.25). When combined with a single propensity-score adjusted NRCS, the estimate was similar, but more precise (0.97, 95% CI 0.62 to 1.53). Four RCTs provided an imprecise estimate of effect of CPAP on risk of stroke (OR 0.96, 95% CI 0.69 to 1.33). Three RCTs were highly imprecise regarding risk of transient ischemic attack (OR 0.95, 95% CI 0.31 to 2.87). Four RCTs yielded an imprecise estimate of the effect of CPAP on risk of myocardial infarction (OR 1.06, 95% CI 0.72 to 1.56).

Six RCTs and 3 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 3 NRCSs found strong associations between CPAP use and reduced risk of (variable) composite CV outcomes (with approximate effect sizes ranging from 0.37 to 0.83). Compliance in the ITT analyses ranged from about 38 to 60 percent. Across studies, the summary effect sizes of ITT and as-treated analyses were not significantly different, although studies that reported both analyses all found nominally stronger effect sizes among compliant CPAP users. The one RCT that reported a within-study analysis reported no association between degree of CPAP compliance and CV outcomes.

Two RCTs reported imprecise effects of CPAP on traffic or home accidents. Two RCTs found no significant difference in risk of incident diabetes; addition of an adjusted NRCS does not alter the conclusion. Between three and five studies (for each analysis) found small, but clinically nonsignificant improvements with CPAP in depression and anxiety symptom scores, measures of executive cognitive function, and QoL. The evidence for other outcomes was sparse, imprecise, and/or inconsistent, including for transient ischemic attack, angina, coronary artery revascularization, congestive heart failure, atrial fibrillation, incident hypertension, functional status, sexual function, and days of work missed.

Subgroup and cross-study comparisons did not elucidate potential modifiers of effect of CPAP. 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 compliant 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 with noncompliant CPAP users.

CPAP versus other active treatments: Long-term clinical 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. One RCT, each, found imprecise estimates of differences in QoL and sexual function between CPAP and MAD.

Two RCTs found no significant differences in functional status between patients using autoCPAP or fixed CPAP suggesting moderate 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 clinical outcomes. Also, no study explicitly reported analyses of endpoint surrogacy or mediation. None of the clinical event 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 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 clinical outcomes.

Limitations

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. Subgroup analyses to determine which patients may most (or least) benefit from CPAP were rare and inconclusive. Studies rarely reported analyses evaluating association of compliance with CPAP use and clinical 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 compliance). The validity of evaluated QoL, cognitive function, and sexual function is unclear. 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. However, there is insufficient evidence to determine how study results may vary based on the different breathing measures and criteria. Studies have not directly evaluated the validity of changes in breathing or sleepiness measures (e.g., AHI or ESS) as intermediate or surrogate measures of long-term clinical outcomes. Analysis of study-level correlations between changes in potential intermediate or surrogate measures and clinical 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 translate to improvements in clinical outcomes.

Based on RCT data alone, there is low SoE that CPAP use does not affect the risk of all-cause mortality, stroke, myocardial infarction, composite CV outcomes, driving accidents, and incident diabetes. These conclusions were in large part based on imprecise, nonsignificant effect

sizes. There is low SoE that CPAP does not yield clinically meaningful changes in depression and anxiety symptoms, cognitive function, or QoL. Adjusted NRCSs were concordant with the RCTs in terms of the direction of associations with long-term clinical outcomes. Combining RCT and NRCS evidence yielded a suggestion that CPAP reduces the risks of all-cause mortality (low SoE). This conclusion may be most applicable to older adults (at increased risk of dying) and longer-term followup. NRCS evidence did not change conclusions about other outcomes.

It is unclear whether the lack of significant differences between compliant 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 compliant and noncompliant 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 (maybe such as increased communication with the sleep clinic) or that even the "low dose" of CPAP achieved by noncompliant users is effective. However, such explanations are just conjectures and would need to be explored critically.. There is also concern that many of the NRCSs, particularly those that conducted as-treated analyses, may be subject to inherent, incompletely adjusted-for, biases related to self-selection of who chooses to (or is chosen for) use of CPAP and is compliant with using the device.

For the various outcomes, 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. It would not be unexpected for future evidence to alter these conclusions.

Two small RCTs provide low SoE that any changes in depression and anxiety symptoms are not affected by the choice of CPAP versus MAD. Two small RCTs provide moderate SoE that a measure of functional status is not affected by the choice of autoCPAP versus fixed CPAP. The moderate SoE suggests that we have moderate confidence that the summary estimates (and their confidence intervals) are close to the true effect. The body of evidence has some deficiencies. Additional evidence would most likely support the current findings, but some doubt remains.

The effect of CPAP on most long-term clinical outcomes is unclear, due to insufficient evidence from sparse studies and/or highly imprecise estimates. Additional studies are needed before we have a clear understanding of the potential effects of CPAP on long-term outcomes for patients with OSA. 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 clinical outcomes.

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 a combination of the two (mixed). The first of these, obstructive sleep apnea-hypopnea syndrome more commonly called obstructive sleep apnea (OSA), is the most common type of sleep apnea.[1]

Definition and Severity of Obstructive Sleep Apnea

Sleep apnea is primarily diagnosed with sleep tests that measure sleep time (and often other sleep measures) and respiratory events. OSA is distinguished from central sleep apnea by the presence of respiratory effort during episodes of apnea and hypopnea (in central sleep apnea, respiratory effort is lacking). In OSA, the volume of airflow is diminished despite appropriate ventilator effort. It differs from central sleep apnea, in which there is a reduced drive to breathe, reduced ventilator 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 including oxygen desaturation index (ODI), respiratory disturbance index (RDI), 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 its scoring manual in 1999 (known as the "Chicago" Criteria).[2] They have amended their definitions of breathing events, sleep time, and how these are measured multiple times since their first set of criteria, with major revisions in 2007[3] and 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, respiratory effort related arousals (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 "respiratory disturbance index" (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 [HTN], 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 for use during sleep.[8] The CPAP machine directly relieves the obstruction by counteracting airway narrowing through the delivery of compressed air (under pressure) to the oropharynx, thereby splinting the airway (keeping it open with increased air pressure). 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 HTN, ischemic heart disease, or history of stroke."[9] Of note, in contrast with how apneas and hypopneas are defined in at least some studies, CMS requires that "Apnea is defined as a [100%] cessation of airflow…[and h]ypopnea is defined as an abnormal respiratory event… with at least a 30% reduction in …airflow... with at least a 4% oxygen desaturation." The 2001 CMS Coverage Decision Memorandum for CPAP, which had substantively the same criteria as current CMS policy, noted that their criteria were derived from the inclusion criteria of studies of CPAP devices and that there was not any direct evidence to support the use of the criteria.[10]

Non-CPAP treatments 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.

For specific groups of patients, other interventions that are used in clinical practice for treatment include atrial overdrive pacing for patients with nocturnal bradycardia, weight loss interventions (including bariatric surgery), and various surgical interventions that aim to alter the anatomy of the air passages to alleviate postulated obstructive mechanisms. These specialized

interventions are not first-line treatments, are not a direct comparator to CPAP for the majority of incident patients, and are, thus, 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, 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 outcomes are more rarely reported.[11, 12] As described in the 2011 Agency for Healthcare Research and Quality (AHRQ) Comparative Effectiveness Review on OSA,[12] 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 clinical 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.[13-15] 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 clinical outcomes, as well as whether, and in whom, CPAP may be a clinically effective treatment modality.

Purpose of the Review

CMS nominated the topic to AHRQ for a Technology Assessment. The review provides information to address various background Contextual Questions on diagnostic measures, treatment modalities, and CPAP specifically. It also summarizes evidence on long-term clinical health outcomes with CPAP treatment and assesses the validity of surrogate and intermediate measures (e.g., AHI) for clinically significant outcomes.

Methods

Review Approach

The Brown Evidence-based Practice Center conducted this systematic review (SR) based on the AHRQ Methods Guide for Effectiveness and Comparative Effectiveness Reviews (available at https://effectivehealthcare.ahrq.gov/topics/cer-methods-guide/overview).[16] This SR also reports in accordance with the Preferred Items for Reporting in Systematic Reviews and Meta-Analyses (PRISMA),[17] A Measurement Tool to Assess Systematic Reviews (AMSTAR 2),[18] 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 Agency for Healthcare Research and Quality's (AHRQ) 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 *clinically significant outcomes*?

> **KQ 1a:** How are respiratory disturbance events (apnea, hypopnea, arousal) defined in each study? What are the diagnostic criteria for OSA (or criteria to treat with CPAP) in each study? How do the diagnostic criteria relate to time of the American Academy of Sleep Medicine (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 clinically significant outcomes?*
>
> **KQ 1c**: Do the clinical effects or harms of specific CPAP devices differ by patient subgroups, duration of followup, or particular CPAP features?
>
> **KQ 1d**: Summarize the methodological issues in the existing studies.

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

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

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

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 clinically significant. Specifically, this analytic framework depicts the chain of logic that evidence must support to link the interventions to improved health outcomes. The figure lays out which 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.[12, 19-21] For more recent articles, *de novo* literature searches were conducted in Medline (via PubMed), Embase, 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 November 18, 2019. We also searched the ECRI guidelines Trust[22] for relevant guidelines published in the last 5 years and the US Food and Drug Administration (FDA) medical device databases for summaries of safety and effectiveness that may include study results not published elsewhere.[23]

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.

Searches will be updated during the peer and public posting period.

Study Selection

Table 1 presents the major eligibility criteria for each KQ. More detailed criteria are presented in Appendix A.

Table 1. Study eligibility criteria, per Key Question

Eligibility Categories	Criteria
Both KQs: Population*	Adults (≥18 years) *Exclude studies with any pregnant women* *Exclude studies in which any participants are reported to have, at baseline, central sleep apnea (from any cause including prior stroke, severe heart failure, among others), obesity hypoventilation syndrome (Pickwickian syndrome), neuromuscular disease, Parkinson disease, Down syndrome, Prader-Willi syndrome, major congenital skeletal abnormalities, narcolepsy, narcotic addiction, Alzheimer disease, epilepsy and or with mild cognitive impairment*
Both KQs: Intervention/Comparator*	*Exclude studies of surgical interventions for sleep apnea or bariatric surgery*
Both KQs: Outcomes*	Major clinical outcomes • Death • Cardiovascular and cerebrovascular events or incident diagnosis • Motor vehicle accidents • Composite outcomes that include only major clinical outcomes (e.g., major adverse cardiovascular events defined as including all-cause mortality) Other patient-centered and/or clinically significant outcomes • Other cardiovascular outcomes • Objective measures of cardiovascular severity (categorized, not continuous measures such as intima media thickness) • Incident hypertension (or regression to normotension) • 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)*	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
KQ 1 (CPAP treatment): Population †	Obstructive sleep apnea (as per study criteria)

Eligibility Categories	Criteria
KQ 1: Intervention †	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 compliance/adherence (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 †	No CPAP Allow assignment to no CPAP, lack of prescription for CPAP, lack of use of CPAP (e.g., due to noncompliance/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 †	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) †	As listed above, for both KQs
KQ 1: Design †	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 †	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 clinical outcome and other studies that reported sufficient data to analyze a potential association between the change in the measure and the clinical outcome. Studies had to compare two groups (either two interventions or two groups based on participant characteristics, like gender).
KQ 2: Population †	As listed above, for both KQs
KQ 2: Intermediate/surrogate measures †,‡	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)*
KQ 2: Outcomes †	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

Eligibility Categories	Criteria
KQ 2: Timing (minimum followup duration) †	As listed above, for both KQs
KQ 2: Design †	Comparative studies informing on validity or person-level associations of change in sleep and breathing measure(s) with outcome(s) 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.

* See sections below for eligibility criteria specific to either KQ 1 or 2.
† These criteria are in addition to those listed for "Both KQs".
‡ Variables of interest evaluated regarding their association with clinical 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 clinical outcomes) and KQ 2 (on correlation between intermediate/surrogate measures and clinical 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.

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 randomized controlled trials (RCTs), we used all the items from Cochrane Risk of Bias 2.0 tool,[24] 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 nonrandomized comparative studies (NRCS), we included assessments of specific elements from the Risk Of Bias In Non-randomised Studies – of Interventions (ROBINS-I) tool,[25] 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).

We 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 nonbaseline variables were included, and the comparability of baseline characteristics between groups after propensity score adjustment.[26] 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. 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,[27] but also allows for an assumption that NRCS evidence is inherently flawed and should be omitted from synthesis. 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 RCT comparing CPAP with no CPAP, we calculated the number-needed-to-treat (NNT) or to harm (NNH) based off of 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.[28] 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 clinical 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 clinical 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.[29] 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 (CVD), cerebrovascular disease (CeVD), 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) 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.

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.[16, 30] 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). 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, 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. However, we determined a SoE rating for only RCTs alone and, separately, for all studies. We did not determine a SoE rating for NRCS 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 clinical 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. For KQ 1, 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, and 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 or biometric measures should be used for diagnosis, severity assessment, prognosis and treatment monitoring.[31, 32]

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.

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).[33] It is believed 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, apnea index (AI), respiratory disturbance index (RDI), oxygen desaturation index (ODI) , respiratory effort related arousals (RERA), hypopnea index and others quantify *number of respiratory events per unit time.* As such, they can be confounded by sleep duration. In contrast, metrics of cumulative exposure to reduced inhaled air during a night of sleep (e.g., total number of events, total duration of events by type, time spent with oxygen saturation below a specific threshold) are not related to, and are thus unconfounded by, sleep duration.

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 obstructive sleep apnea (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.[34] 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.[35]

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,[36] by the morphology of the oxygen desaturation events,[37]or by 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.[36, 37] Generally, AHI is not well-correlated with symptoms and signs of OSA.[38] 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).[39, 40] 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).[41]

In 1999, an American Academy of Sleep Medicine (AASM) Task Force recommended counting only hypopneas accompanied by a 3 percent oxygen desaturation or a respiratory event-related arousal (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.[42] 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 2018 AASM update dropped this second definition of hypopnea.[42] 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 clinical 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% airflow reduction and 4% desaturation) versus 13.4 (95% CI 6.8 to 24.1) with the 2012/2018 AASM definition (at least 30% airflow reduction plus ≥3% desaturation and/or arousal).[43] 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 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.[44] 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,[45] sleep labs,[46] the general population,[47] and in patients with chronic diseases to screen for OSA.[48] 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,[49] various scoring algorithms have been proposed that do not count all items equally.[45] 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[44] (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 clinical 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).[50, 51] 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.[51]

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.[52-55] The ESS exhibits moderate or weaker correlations with other constructs (e.g., maintenance of wakefulness test, multiple sleep latency test, AHI and other measurements).[53] In all, we consider it analytically valid.[35] However, ESS has limited clinical validity in detecting various definitions of OSA.[12, 56] 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][57], EuroQol-5D) and specialized (e.g., Calgary Sleep Apnea Quality of Life Index [SAQLI],[58] Functional Outcomes Sleep Questionnaire [FOSQ][59]) health status tools, health indices, and QoL tools have been used in sleep apnea assessments.[60] 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 quality of life (hrQoL). There is a wide range of definitions of hrQoL.[61] As defined by the U.S. Food and Drug Administration, hrQoL is a "multidomain concept that represents the patient's general perception of the effect of illness and treatment on physical, psychological, and social aspects of life".[62] Alternatively, hrQoL is "the value assigned to duration of life as modified by the impairments, functional states, perceptions, and social opportunities that are influenced by disease, injury, treatment, or policy".[63]

We briefly comment on FOSQ, a disease-specific health status instrument that measures impact of sleepiness on activities of daily living.[59]

FOSQ is a self-administered tool that 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]);[64] although we found no studies that have validated it against established OSA-related hrQoL 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]).[65] It was found to be valid across several validity domains, including responsiveness to CPAP use and compliance. There is also some information that FOSQ has limited clinical validity in detecting various definitions of OSA (i.e., screening).[12, 56] 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.[12, 66, 67] They help reposition the lower jaw to alleviate upper airway obstruction during sleep. For some patients with specific anatomic variations, surgery may be indicated,[68] including uvulopalotopharyngoplasty,[69] maxillomandibular advancement,[70] and adenotonsillectomy (particularly in children). These surgeries correct structural airflow obstructions and, thus, increase the volume of inhaled air during sleep.

Newer treatments include nasal expiratory positive airway pressure (EPAP) devices,[71] oral pressure therapy,[72] and implantable hypoglossal nerve stimulation devices.[73, 74] 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.[71, 75]

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 mouth piece 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.[72]

Hypoglossal nerve stimulation by means of implantable devices is being examined for patients with moderate to severe OSA.[73, 74, 76] 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 can 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."[77] 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[78] 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 between 1976 and 1994.

For the 163 Premarket Notification records on CPAP devices, we extracted an identification number, the device name and manufacturer, and the approval date. 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. The extracted data for the 163 records is presented in Appendix Table D-4.

For the 143 CPAP device records approved in 1996 or later, we found no 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 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.

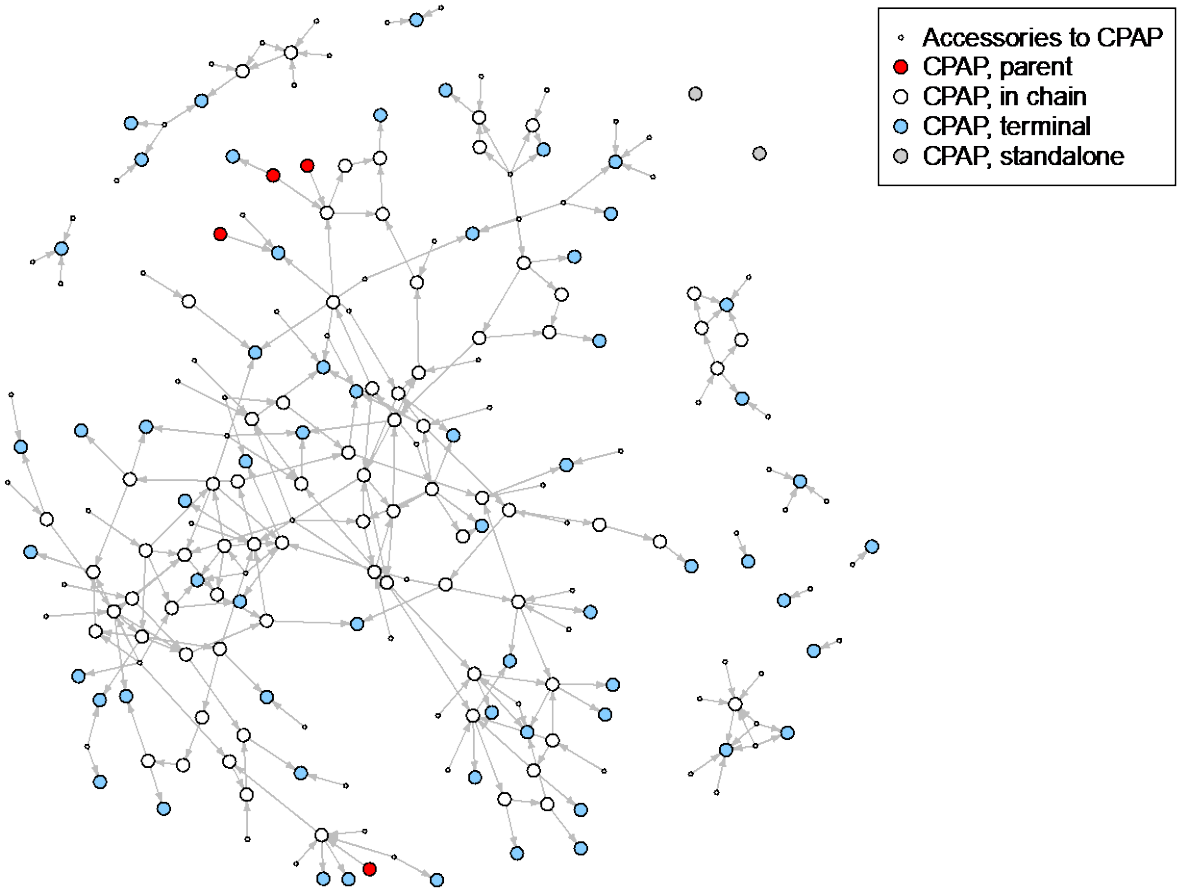
We found a total of 266 citation relationships. Figure CQ3 graphs the citation relationships reported in the 143 device records. The figure demonstrates 13 groups of records that are connected with citation relationships. The largest relationship group includes 169 device records. The other relationship 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 CQ3) 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 CQ3. Citations between records to support equivalence claims



Shown are citation relationships (266 grey arrows) between 143 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. Bi-level machines (BiPAP) 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 BiPAP machines) 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. We did not find a description of the objective function the devices try to optimize, which feedback signals they use, or the integration of the feedback signals. These machines are often referred to as 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 machines marketed for patients with either obstructive or mixed/central sleep apneas.

Humidity Control

Several machines use heated humidification that may help reduce nasal congestion or dryness, and thus improve adherence to use.

Monitoring and Documentation of Use

Machines 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 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 machine (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 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.

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, 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. 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,[79, 80] hypertension,[81] impaired cognitive function (and Alzheimer disease),[82] type 2 diabetes,[83] anxiety and depression symptoms,[84] impaired quality of life,[85] and impaired sexual function in both men and women.[86, 87] 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 clinical outcome goals. In practice, machines 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 15,333 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 47 primary studies (reported in 70 articles) were eligible and included. The reasons for exclusion of articles were: duplicate analysis (no unique data) (n=584), KQ 2: no apnea-hypopnea index (AHI) change data (n=258), no outcome of interest (n=186), too short followup (n=145), SR (n=88), no results given (n=78), full text not available (n=52), KQ 2 N<30 total (n=29), excluded population (n=28), not primary study (or SR) (n=17), no comparator of interest (KQ 1) (n=13), KQ 1 not continuous positive airway pressure (CPAP) device (n=11), KQ 1 N<10/arm (n=10), not obstructive sleep apnea (OSA) (n=9), case-control/cross-sectional study (n=7), nonrandomized comparative study (NRCS) without multivariable analysis (KQ 1) (n=6), and conference abstract (n=2). See 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 polysomnography 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 polysomnography measures used.

Findings

As displayed in Table 2, we categorized each of the 47 eligible studies (for both KQs) 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 polygraphy) monitors used. Several issues are evident.

The majority of studies did not explicitly report full criteria or definitions; one-third (16/43, 37%) of studies that evaluated the apnea-hypopnea index (AHI) (excluding studies that evaluated only oxygen desaturation index [ODI] or were registry studies without AHI data) omitted any explicit apnea or hypopnea definitions; only 17 of 43 (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 43 studies that reported at least some apnea or hypopnea definitions, 11 defined apnea as 100 percent airflow cessation, but 7 used lower thresholds down to 75 percent (well within the definition of hypopnea used by most

studies). Among the 17 studies that explicitly reported hypopnea criteria (and used a single hypopnea definition), about half (8/17) 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 32 studies that explicitly reported oxygen desaturation thresholds (either as part of the hypopnea definition or to determine ODI), a bit more than half (18/32, 56%) used a 4 percent desaturation threshold and about one third (12/32, 38%) 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 29 studies that cited published criteria to define apneas or hypopneas, 25 (86%) cited some version of the American Academy of Sleep Medicine (AASM) criteria. However, among the 16 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 clinical outcomes) differed in relation to how apnea, hypopnea, ODI, or OSA were defined. For most clinical 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*	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 †	NR	100%	50%	4%	n/a	AASM 1997
Tegelberg 2012 23620682	Cohort	III home	100%	50%	4%	n/a	AASM 1999
Crawford-Achour 2015 25700873	*NRCS CPAP vs No*	*III home*	*100%*	*50%*	*3%*	*3%*	*NR*
Nakamura 2009 No PMID	*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
Wu 2015 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 ‡	I lab or III home	100%	NR	3%	n/a	Other
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 ‡	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 Other*
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 †	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
de Ruiter 2018 28913630	RCT Other ‡	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 ‡	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 #	NR	NR	NR	NR	n/a	AASM 1999

Study PMID	Design*	Monitor Type	Apnea Threshold	Hypopnea Threshold	Hypopnea Desaturation %	ODI Desaturation %	Criteria Cited
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*
de Vries 2019 31596213	RCT CPAP vs. Other #	I lab or II home	NR	NR	NR	n/a	AASM No date
Fernández-Julián 2017 29152745	NRCS Other ‡	II home	NR	NR	NR	n/a	AASM 2012
Jara 2018 29800001	*NRCS CPAP vs No*	*I lab*	*NR*	*NR*	*NR*	*NR*	*AASM 2015*
Kingshott 2000 10712335	Cohort	I lab or II home	NR	NR	NR	NR	NR
Kushida 2011 21804670	RCT CPAP vs. CPAP †	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
Wu 2016 26993342	**RCT CPAP vs No**	**I lab or II home**	**NR**	**NR**	**NR**	**NR**	**AASM 2007**
Jennum 2015 25914563	*NRCS CPAP vs No*	*Any*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*

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 darker/greyer shades indicating more recent years. Other (non-AASM) 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.

* 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.
† Or other positive airway pressure, such as auto-adjustable.
‡ Not a comparison of CPAP.
# E.g., CPAP vs. mandibular advancement device

Key Question 1: CPAP Versus No CPAP Treatment Effect

Key Points

- 25 studies compared CPAP with no CPAP. These included 12 RCTs (N = 6019, total) and 13 NRCSs (N = 32,062; 25,389 in one study that reported only on all-cause mortality, 6673 in other NRCSs); 14 intention-to-treat (ITT) analyses and 11 as-treated analyses. Two RCTs used sham CPAP as the comparator. No study reported within-study correlations among outcomes (e.g., effect on AHI and effect on all-cause mortality).
- Few studies were explicitly powered for long-term clinical outcomes: 2 RCTs were designed to be powered for composite CV outcomes (SAVE, RICCADSA); however, RICCADSA was ultimately underpowered. One RCT (APPLES) was powered for neurocognitive outcomes. No study was explicitly powered for any other outcome. 14 studies (56%) reported intention-to-treat analyses; 11 (44%) conducted as-treated analyses of compliant CPAP users. The as-treated analyses may not have been able to fully account for possible biases related to self-selection regarding CPAP use and compliance.
- 3 RCTs provide an imprecise estimate of whether CPAP alters the risk of all-cause mortality (thus, a conclusion of low SoE of no effect). Additional evidence from 5 adjusted NRCSs were consistent with the RCTs in direction of association, but were statistically significant. Compared with the RCTs, the NRCSs had higher death rates related to older age (in 2 NRCSs) and longer-term followup (up to 11 years). Combining the RCTs and NRCSs provided low SoE that CPAP reduces the risk of all-cause mortality; this conclusion may be most applicable to older adults and longer-term followup. Omitting the one very large NRCS did not alter the conclusion. The ITT analyzed studies (in which compliance ranged from 38% to 60%) found similar associations as the as-treated studies (of compliant users). No definitive explanation was provided for the lack of difference in effect between analyses with and without noncompliant CPAP users.
- 3 RCTs provide insufficient evidence regarding the effect of CPAP on risk of cardiovascular (CV) mortality, but together with 1 NRCS, there is low SoE that CPAP does not affect risk of CV death.
- 4 RCTs provide low SoE that CPAP does not affect risk of stroke or acute myocardial infarction (MI).
- 6 RCTs provide mostly imprecise estimates of whether CPAP alters the risk of various composite CV outcomes (thus, a conclusion of low SoE of no effect). Additional evidence from 3 adjusted NRCSs did not change this conclusion. Across studies, the as-treated analyses of CPAP users did not find significantly different associations than the ITT analyses of assignment to CPAP use (in which compliance ranged from about 38% to 64%).
- 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 atrial fibrillation (AFib). For each outcome, there were sparse studies and/or effect estimates were highly imprecise.

- 2 RCTs provide low SoE that CPAP does not affect the likelihood of driving accidents after 1 or about 3 to 4 years of use.
- 2 RCTs found no evidence of an effect of CPAP on risk of type 2 diabetes (DM) (low SoE); inclusion of 1 additional NRCS does not alter the conclusion.
- There is low SoE that CPAP does not result in a clinically significant improvement in depression symptoms (based on 4 RCTs) and anxiety symptoms (3 RCTs), but the differences compared with no CPAP were statistically significant.
- 3 RCTs found that CPAP results in better cognitive function as measured on two standard tests (low SoE), but the differences compared with no CPAP use were small and not clinically significant. 1 additional NRCS did not alter the conclusion.
- 8 RCTs found that CPAP results in better quality of life (QoL) and functional status scores, as measured on standard tests (low SoE), but the differences compared with no CPAP use were small and not clinically significant. 1 additional NRCS did not alter the conclusion.
- There is insufficient evidence regarding the effect of CPAP on other outcomes, including hypertension, 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.
- Based on the evaluated long-term, comparative studies, there is insufficient evidence regarding the risk of adverse events specific related to CPAP use.
- Across outcomes, there was no (or limited) evidence regarding possible heterogeneity of treatment effect (different effects in different groups of patients). The studies did not provide evidence that CPAP use is more (or less) effective in any specific subgroup.

Evidence Base

Twenty-five studies (in 47 publications) compared CPAP with no (or rarely sham) CPAP treatment. These included 12 RCTs[13, 88-98] and 13 NRCS[99-114] with multivariable adjustments for outcomes of interest. The 12 RCTs included a total of 6019 participants. One of the NRCSs was relatively large (N = 25,389); the remaining 12 NRCs included 6673 participants. Among the 13 NRCSs, three 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 10 NRCSs used traditional logistic regression multivariable analyses to control for specific possible confounders. However, it is unclear whether the NRCSs, particularly those that reported as-treated analyses, could have fully accounted for possible biases related to self-selection regarding CPAP use and compliance.

Study eligibility criteria were variable across studies. Among the 12 RCTs, four included adults with OSA regardless of other major comorbidities,[90, 91, 93, 97] four restricted primarily to participants with CVD or CeVD (one post-revascularization),[13, 92, 94, 98] three excluded patients with cardiovascular disease (CVD),[88, 89, 96] and one restricted to patients with type 2 DM.[95] One RCT was restricted to older adults (≥65 years)[93] and six trials excluded older patients (<65-75 years).[13, 89, 90, 92, 96, 98] All but three of the 13 NRCSs did not restrict their participant samples by comorbidities: one was restricted to patients who recently had percutaneous coronary intervention with a drug-eluting stent,[105] one excluded patients with CVD or cerebrovascular disease (CeVD) or CVD risk factors,[107] and one excluded patients with DM.[103] Among the NRCSs, three were restricted to older adults (≥60, 65, or 80 years).[107-109] Most of the 12 RCTs

included patients with at least "moderate" OSA (e.g., AHI ≥15); two RCTs restricted to patients with "milder" OSA (AHI 10-30 or 5-45)[88, 97] and a single study restricted to patients with "severe" OSA (ODI ≥12, which was determined to be equivalent to AHI ≥30).[13]

Fourteen of the studies reported ITT analyses comparing participants who were prescribed CPAP with those not prescribed CPAP (or given a sham CPAP device). The remaining 11 studies conducted as-treated analyses comparing compliant CPAP users mostly with all nonusers, including both those who never started CPAP and those who were noncompliant. Two RCTs gave the control participants sham CPAP devices.

Each study is described in some detail, including study design, analysis type (ITT vs. as-treated), 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, compliance with CPAP use varied across studies. We were unable to discern any obvious patterns across studies regarding low versus high lack of compliance, including duration of the studies. Among RCTs that reported on compliance, four reported compliance rates of 38 to 43 percent (at 6 months in three RCTs, and across an average of 44 months in the SAVE trial),[13, 90, 91, 98, 115] and three reported compliance rates of 60 to 64 percent (at 6 months and 1 and 4 years).[89, 94, 95] Average CPAP usage varied from 2.5 to 5.8 hours per night.

Of note, in regards to Key Question 1b (within-study concordance between apnea and hypopnea indices, sleep questionnaires, and clinical 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.

Mortality and Cardiovascular Outcomes

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.

Randomized Controlled Trials

Barbé 2012 randomly allocated relatively young, otherwise healthy participants to CPAP (n=357) or no CPAP (n=366) for 4 years.[89] 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 (Epworth Sleepiness Scale [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 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 hypertension (HTN), which was not an outcome of interest for the current report. At 4 years, 64 percent of CPAP users were compliant (≥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 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.[92] However, participants who were allocated to CPAP but had poor compliance (mean CPAP use <4 hours/night) were excluded from analysis. This applied to 9.5 percent (4/42) of participants randomized to CPAP. Thus, the study was analyzed based on "as-treated" groups. Participants were included if they had both 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 body mass index (BMI) of 27.7 and a mean neck circumference of 41.1 cm. Ten percent of patients assigned to CPAP were noncompliant and were excluded from analysis. The study was not explicitly powered for any clinical 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).

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.[116] The trial was conducted and analyzed with an ITT approach; participants were not excluded from analysis based on CPAP compliance. The complete MOSAIC trial included 391 participants.[90] All 391 participants in the original trial were followed for 6 months while 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 sleepiness symptoms (ESS>9). 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. Compliance (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.[93] The trial was conducted and analyzed with an ITT approach; participants were not excluded from analysis based on CPAP compliance. 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)%, heart failure (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. The study did not report compliance information. 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.[94] The trial was conducted and analyzed with an ITT approach; participants were not excluded from analysis based on CPAP compliance. 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.[13, 115, 117-119] The trial was conducted and analyzed with an ITT approach; participants were not excluded from analysis based on CPAP compliance. However, to be eligible, patients had to be compliant (≥3 hr/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 (ODI ≥12) and had a diagnosis of CAD or CeVD. The study was designed to be enriched in patients at increased risk for CV disease 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 compliance, 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, heart failure, or transient ischemic attack [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 compliant 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 OR results within meta-

analyses). They also reported several sensitivity analyses using other analytic techniques. We also summarize their "as-treated" analyses of those compliant with CPAP therapy (≥4 hours/night for first 2 years). This analysis was conducted as a propensity-score matching one-to-one of compliant 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 CV disease, 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, hypertension, heart failure, 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) and outcomes.

Jennum 2015 was an adjusted NRCS comparing participants in the Danish National Patient Registry diagnosed with OSA who used CPAP (n=13,631) with those not using CPAP, including those who were noncompliant with CPAP (n=11,758) for 3 years.[111] Thus, this study conducted an "as-treated" analysis based on actual use of CPAP. Participants were included if they were diagnosed with OSA, by ICD-10 code. 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, neck circumference and preexisting diseases were provided. 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.[114] The study implicitly conducted an ITT analysis, as it was based on prescription, not use, of CPAP. Participants were included if they were over 40 years and able and willing to undergo a home polysomonography. 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. After propensity score matching, 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 was rated as low risk of bias based on use of propensity score analysis and generally complete reporting. No flaws were noted in their propensity score analysis.

López-Padilla 2016 reported an adjusted NRCS of older adults (≥80 years) with OSA of CPAP use.[109] The study conducted an "as-treated" analysis of compliant CPAP users (≥4 hours/night, n=79) versus not prescribed (n=76). 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 hypertension (78%), cardiac failure (35%), tobacco use (35%), DM (30%), hyperlipidemia (28%), arrhythmias (26%), ischemic heart disease and stroke (15% each). The

study was rated as moderate risk of bias, primarily based on their use of a regression model for adjustment of baseline confounders.

Myllylä 2019 was an adjusted NRCS evaluating patients who had been prescribed CPAP for OSA.[113] 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 6 years. Thus, this study conducted an as-treated analysis 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%). 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 2009 reported an adjusted NRCS, based on Okinawa Nakamura Sleep (**ONSLEEP**) registry, comparing participants who started CPAP (n=1137) with those not starting CPAP (n=1137) at 7-year followup.[101, 102] The study implicitly used a "quasi-ITT" approach in that all patients who agreed to use CPAP (and used it for at least 1 night) were evaluated as CPAP users, regardless of compliance; however, CPAP nonusers were either not prescribed CPAP or did not commence CPAP use despite a prescription for use. Participants were included if they were diagnosed with OSA (AHI>5). Among 4000 participants in the registry, the mean age was 51 years, 80 percent were male, and the mean BMI was 27.9. The article did not report baseline comorbidities, including CVD. A conference abstract in 2019 reported a propensity score-matched analyses of long-term all-cause mortality (about 6.5 year followup), but not for other outcomes. Few details were reported. A full-text manuscript of this analysis is in press and we expect it will be available for inclusion during our literature search update. For all-cause and CV death, the study was tentatively rated as low risk of bias based on use of propensity score analysis, pending more-complete evaluation of their methods.

Ou 2015 was an adjusted NRCS comparing participants who used CPAP (n=36) with those not using CPAP (n=88) for 5 years.[108] Similar to Nakamura 2009, the study implicitly used a quasi-ITT approach in that all patients were offered CPAP and the analysis compared those who agreed to start CPAP (regardless of compliance) 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 compliant (4 hours/night, ≥5 nights/week) Participants in the two arms had similar prevalence in preexisting diseases, including coronary artery disease (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.

Schipper 2017 was an adjusted NRCS comparing participants who used CPAP (n=162) for at least 4 hours per night with those not using CPAP at least 4 hours per night (n=121) for 6 years.[110] Thus, this study conducted an as-treated 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 moderate risk of bias, primarily based on their use of a regression model for adjustment of baseline confounders.

Wu 2015 was an adjusted NRCS comparing participants who used and were compliant with CPAP (≥4 hours/night, ≥70% of nights, >3 months; n=128) with those who refused or were noncompliant with CPAP (n=167) for 5 years.[105] Thus, this study conducted an as-treated analysis based on actual use of CPAP. Participants were included if they had a percutaneous coronary intervention with a drug-eluting stent for coronary artery disease and had "moderate to severe" OSA (AHI ≥15). 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 moderate risk of bias, primarily based on their use of a regression model for adjustment of baseline confounders.

All-Cause Mortality

Eight studies (3 RCTs and 5 NRCSs) comparing CPAP and no CPAP reported on all-cause mortality (Table CVD.01).[13, 94, 99-102, 108, 109, 111, 114, 116] 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).

Randomized Controlled Trials (All-Cause Mortality)

The three RCTs evaluated different, but overlapping, sets of patients at 2 to 5 years of followup. The MOSAIC trial included otherwise healthy adults 45 to 75 years old with at least "moderate" OSA (ODI ≥7.5). The RICCADSA trial 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). None of the trials was explicitly powered for all-cause mortality. All were analyzed as ITT for CPAP versus no CPAP prescription. Compliance (use of CPAP for ≥4 hours per night) varied across studies at about 40 percent in the MOSAIC and SAVE trials and 60 percent in the RICCADSA trial (see Table CVD.01).

In all RCTs, patients and clinicians were not blinded and only the SAVE trial reported that outcome assessors were blinded. The SAVE trial was, thus, rated to be at moderate risk of bias, but the other two trials were further downgraded to high risk of bias due to high dropout rate (MOSAIC) or crossover rate (RICCADSA).

Across RCTs, effect sizes ranged from 0.33 to 0.91, all statistically nonsignificant. The number-needed-to-treat (NNT) to prevent one death ranged from 47 to 426 across studies, which

corresponds to the range of effect sizes across studies. The summary odds ratio (OR) across the RCTs was nonsignificant at 0.87 (95% confidence interval [CI] 0.59 to 1.29) (Figure CVD.01). Under an assumption of a control rate (risk of death without CPAP use) of 4.3 percent (the meta-analyzed control rate across RCTs), the summary NNT was 186 (95% CI 58 to −85 [i.e., a number-needed-to-harm of 85]). The effect sizes did not correlate with compliance rates in the ITT analyses. SAVE also reported a propensity-score matched analysis of adherent CPAP users, with a still-nonsignificant, but nominally stronger hazard ratio (HR): 060 (95% CI 0.32 to 1.10).

Adjusted Nonrandomized Comparative Studies (All-Cause Mortality)

The five adjusted NRCSs reported on all-cause mortality at a wide range of followup times from 3 to 11 years, overlapping the RCTs but with 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). None of the NRCSs was explicitly powered for any outcome. One NRCS (Lisan 2019) was evaluated as an ITT (CPAP versus no CPAP prescription), two (Ou 2015, Nakamura 2009) reported "quasi-ITT" analyses (all were offered CPAP, but those who agreed to CPAP were compared with those who refused), and two (Jennum 2015 and López-Padilla 2016) evaluated "as-treated" (based on compliant use of CPAP). Ou 2015 reported a 67 percent compliance rate, but the other two ITT analyses did not gather compliance data.

Two of the five NRCSs were deemed to be of low risk of bias because they used appropriate propensity score analyses and, overall, described their studies well (Lisan 2019 and Nakamura 2009). López-Padilla 2016 was rated as at moderate risk of bias related to use of simple regression modeling only. The other two NRCSs were deemed to be of high risk of bias due to inadequate descriptions of the studies, possible selective outcome reporting, and, in one instance, inadequate adjustment (Ou 2015).

Across the NRCSs, adjusted HRs ranged from 0.08 to 0.67, all favoring CPAP and all statistically significant. The summary adjusted HR across the peer-reviewed NRCSs was statistically significant at 0.64 (95% CI 0.56 to 0.73), without statistical heterogeneity (Figure CVD.01). Although Ou 2015 found a substantially stronger (smaller) HR, because of its imprecision, it added very little weight to the meta-analyses (<2% meta-analyses of NRCSs or of all studies). The two as-treated analyses (Jennum 2015 and López-Padilla 2016) had a stronger estimate of the association with all-cause mortality than the ITT NRCSs (0.44 vs. 0.61), but the sets of studies did not differ significantly (P=0.79).

All Studies (All-Cause Mortality)

The summary estimate across all studies (RCTs and NRCSs) was 0.66 (95% CI 0.60 to 0.73), with no statistical heterogeneity (Figure CVD.01). *Post hoc*, we noted that, given its large sample size and thus narrow confidence interval, the NRCS by Jennum 2015 (that compared CPAP users with combined nonusers and those noncompliant with CPAP use in a national patient registry) provided a very large weight (80%) of the meta-analysis; thus, the overall summary estimate largely recapitulated this one study. As a sensitivity analysis, excluding this study resulted in a similar summary estimate, suggestive of a stronger effect, however, with a wider confidence interval: 0.62 (95% CI 0.47 to 0.80; $I^2 = 25\%$).

RCTs Versus NRCSs (All-Cause Mortality)

Although the summary HR for the NRCSs was statistically significant while the summary RCT estimate was not, most likely this was largely related to power. 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. One NRCS (Jennum 2015) has a particularly large sample size, which yielded a very precise effect size estimate, but omitting this study did not change summary conclusions. Notably, the summary estimates did not significantly differ between RCTs and NRCSs (P = 0.29). A 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 (including the conference abstracts) 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.

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 2019 was a large study. Note that none of the RCTs reported on subgroup effects.

AHI Subgroups

The Nakamura 2009 NRCS reported adjusted HRs among different subgroups, including by AHI.[101, 102] 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.[111] We calculated these HRs based on their report of a model that included interaction terms between CPAP use and age group.

Body Weight Subgroups

The Nakamura 2009 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.[101]

Pulmonary Function Subgroups

The Nakamura 2009 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.[101]

Across-Study Heterogeneity

Across studies, the three RCTs and five NRCS were statistically homogenous (I^2 =0%), with mostly similar estimates and largely overlapping CIs. However, the small study by Ou 2015 had a considerably larger effect size (smaller HR) than other studies, although with a very wide CI. As noted, the NRCSs were more likely to be statistically significant than the RCTs, related to study power.

Across studies (Figure CVD.01), the summary effect size of the ITT analyses (0.66, 95% CI 0.51 to 0.86) was nearly identical to that of the as-treated analyses (0.64, 95% CI 0.53 to 0.77). Note that the SAVE trial provides information for both the summary ITT and as-treated analyses. It is unclear whether the lack of evidence of a differences between the ITT and as-treated analyses indicate a true lack of effect (as suggested by a lack of "dose effect" of CPAP compliance) or a lack of statistical power to find a statistical difference. The explanation for the lack of difference in effect sized based on compliance is unclear.

Given the lack of heterogeneity across studies, it is not possible to discern cross-study differences in effect of CPAP based on such factors as OSA or AHI definition, comorbidities, or age.

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; 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. 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

Three RCTs provide low SoE that CPAP does not affect all-cause mortality (Table CVD.02). All trials effect sizes favored CPAP (with NNTs ranging from 47 to 426), but individually and by meta-analysis, the RCTs failed to find a significant difference in risk of death (summary OR 0.87, 95% CI 0.58 to 1.29). In aggregate, the studies were of moderate risk of bias and provided an imprecise estimate of effect.

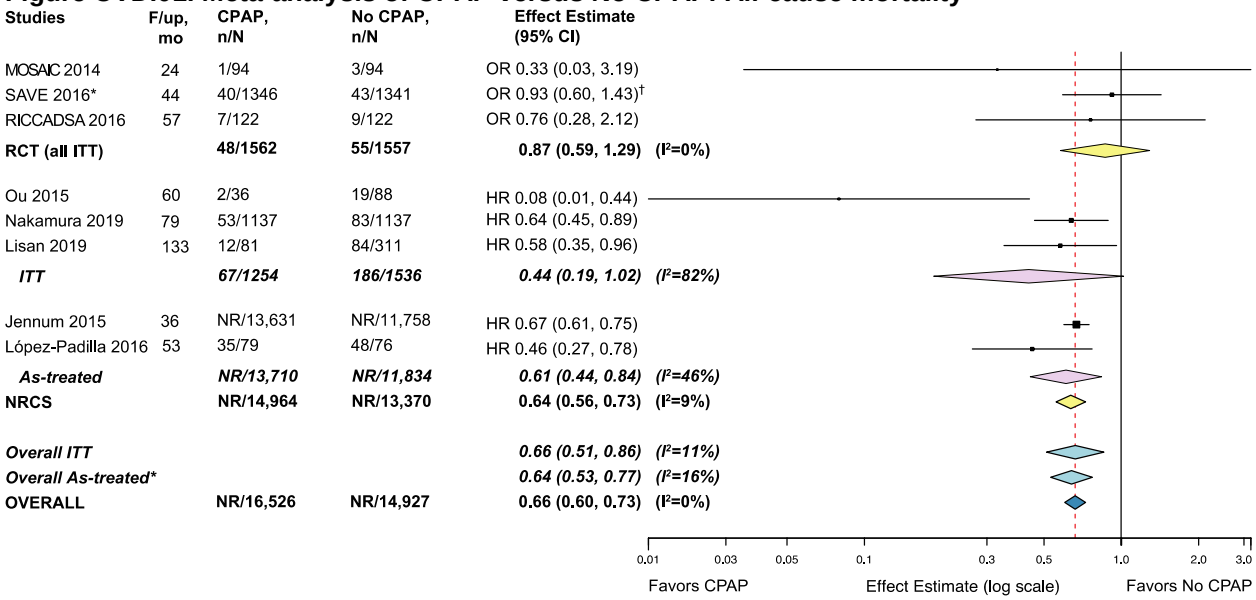
The five NRCSs, one of which had more than 25,000 participants, in aggregate yielded a similar, but somewhat stronger and statistically significant summary effect size as the RCTs (summary effect size 0.62, 95% CI 0.53 to 0.73). Omitting the large NRCS yielded a similar

summary estimate The NRCSs, thus, suggest an association between CPAP use and reduced risk of all-cause mortality.

The difference in conclusions from the RCT evidence and the NRCS evidence is based primarily on a difference in statistical significance; the NRCSs had somewhat stronger effect sizes than the RCTs and the effect sizes were more likely to be statistically significant. However, all studies nominally favored CPAP (effect sizes were all less than 1.0) and the primary difference between RCTs and NRCSs appears to be due most to power considerations. Notably, the statistically significant NRCSs were pertinent to either or both longer-term followup (up to 11 years) and higher risk (related to older age). The NRCSs, particularly those with as-treated analyses, may be biased due to self-selection of CPAP use and compliance, despite attempts to adjust for confounders.

While the RCTs alone do not support that CPAP reduces the risk of all-cause mortality, the similarity in summary estimates between the RCTs and NRCSs (all favoring CPAP) and the consistency in directionality of all studies together provide low SoE that CPAP reduces the risk of all-cause mortality. The reduction in risk may be most applicable to older adults (at increased risk of dying) and longer-term followup. RCTs did not report subgroup analyses. Subgroup analyses were conducted in NRCSs, but there is insufficient evidence to suggest which, if any, participant subgroups in these studies may have had stronger effects. 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. It would not be unexpected for future evidence to alter these conclusions.

Figure CVD.01. Meta-analysis of CPAP versus No CPAP: All-cause mortality



Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, HR = adjusted hazard ratio, I^2 = measure of statistical heterogeneity ranging from 0% (none) to 100%, ITT = intention-to-treat, MOSAIC = Multicentre Obstructive Sleep Apnoea Interventional Cardiovascular trial, NR = not reported, NRCS = nonrandomized comparative study, 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 reported both ITT and CPAP compliant analyses (see Table CVD.01). Thus, the overall ITT and as-treated summary estimates both include the SAVE trial.

† For consistency with other RCTs, OR included in meta-analysis. Using the reported HR instead (see Table CVD.01) resulted in near-identical summary estimates and analyses.

Table CVD.01. CPAP versus no CPAP: All-cause mortality

Study PMID	Design	Analysis (Powered?[1])	Followup Duration (mo)	Arm	CPAP Compliance	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI)[2]
MOSAIC 24508706[116]	RCT	ITT	24	CPAP	38% at 6 mo	1/94 (1.1)	OR 0.33 (0.03, 3.19)*
		(No)		No CPAP		3/94 (3.2)	NNT 47 (16, NNH 50)
SAVE 27571048[13]	RCT	ITT	44	CPAP	42% overall	40/1346 (3.0)	HR 0.91 (0.59, 1.40)[3] NNT 426 (65, NNH 93)
		(No)		No CPAP		43/1341 (3.2)	[Compliant analysis: HR 0.60 (0.32, 1.10)]
RICCADSA 26914592[94]	RCT	ITT	57	CPAP	60% at 1 y	7/122 (5.7)	OR 0.76 (0.28, 2.12)*
		(No)		No CPAP		9/122 (7.4)	NNT 61 (13, NNH 22)
Ou 2015 26068440[108]	NRCS	Quasi-ITT	60	CPAP	67% overall	2/36 (5.6)	***No CPAP vs. CPAP*: Adjusted HR 12.2 (2.28, 64.7)[4]**
		(No)		No CPAP		19/88 (21.6)	
Nakamura 2009[101, 102]	NRCS	Quasi-ITT	79	CPAP	NR	53/1137 (4.7)	**Adjusted HR 0.64 (0.45, 0.89)[5]**
		(No)		No CPAP		83/1137 (7.3)	
Lisan 2019 30973594[114]	NRCS	ITT	133	CPAP	NR	12/81 (14.8)	**Adjusted HR 0.58 (0.35, 0.96)[6]**
		(No)		No CPAP		84/311 (27.0)	
Jennum 2015 25914563[111]	NRCS	As-treated	36	CPAP	100% (as-treated)	NR/13,631	**Adjusted HR 0.67 (0.61, 0.75)[7]**
		(No)		No CPAP		NR/11,758	

[1] Explicitly powered for this outcome?

[2] 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).

[3] Calculated OR = 0.92 (95% CI 0.60, 1.43) used in meta-analysis for consistency with other trials.

[4] Note that this HR is for *no CPAP versus CPAP*, opposite of other comparisons. Adjusted for hypertension, coronary heart disease, diabetes, AHI. We inverted the HR for CPAP vs. no CPAP in the meta-analysis: approximate HR 0.08 (95% CI 0.02, 0.44).

[5] Propensity score analysis: age, sex, lifestyle (smoking, drinking), comorbid conditions including cardiovascular disease and/or stroke, other comorbid conditions, Epworth Sleepiness Scale (ESS) and apnea-hypopnea index (AHI).

[6] Propensity matching (including age, sex, ethnicity, education level, body mass index (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).

[7] Adjusted for confounders without specifying what confounder were exactly.

Study PMID	Design	Analysis (Powered?[1])	Followup Duration (mo)	Arm	CPAP Compliance	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI)[2]
López-Padilla 2016 27198943[109]	NRCS	As-treated	53	CPAP	100% (as-treated)	35/79 (44.3)	**Adjusted HR 0.46 (0.27, 0.78)[8]**
		(No)		No CPAP		48/76 (63.2)	

Statistically significant results are in bold font.

Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, HR = hazard ratio, hr/noc = hours per night, 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, 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.

* Calculated

[8] Adjusted for age, sex, BMI, alcohol intake, diabetes, stroke, and ischemic cardiac disease. Note that study restricted to age ≥80 years old.

Table CVD.02. 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 Estimate (95% CI)
Death, all-cause	RCT	3 (3119)	Moderate	Consistent	Imprecise	Direct	Clin heterog	Low	No evidence of effect OR 0.87 (0.58, 1.29) NNT 186 (**58, −85***)
	adj NRCS	4 (31,328)	Moderate	Consistent	Precise	Direct	Clin heterog	not evaluated	Assn with reduced death 0.66 (0.60, 0.73)
	Overall	*7 (31,298)*	*Moderate*	*Consistent*	*Precise*	*Direct*	*Clin heterog*	*Low*	*CPAP may reduce death* [9] *HR 0.67 (0.61, 0.74)*
Death, cardiovascular	RCT	3 (3654)	High	Consistent	Highly imprecise	Direct	Clin heterog	Insufficient	No conclusion
	adj NRCS	1 (2274)	Low	N/A	Imprecise	Direct	Single study	not evaluated	No conclusion
	Overall	*4 (5928)*	*High*	*Consistent*	*Imprecise*	*Direct*	*Clin heterog*	*Low*	*No evidence of effect* *ES 0.97 (0.62, 1.53)*
Stroke	RCT	4 (3885)	Moderate	Consistent	Imprecise	Direct	Clin heterog	Low	No evidence of effect OR 0.96 (0.69, 1.33) NNT **1161 (149, −142*)**
TIA	RCT	3 (3641)	Moderate	Moderately inconsistent	Imprecise	Direct	Clin heterog	Insufficient	No conclusion
AMI	RCT	4 (3885)	Moderate	Consistent	Imprecise	Direct	Clin heterog	Low	No evidence of effect OR 1.06 (0.72, 1.56) **NNT −685* (145, −74*)**
Angina[10]	RCT	2 (2917)	Moderate	N/A[11]	Variable[12]	Direct	Single study[13]	Insufficient	No conclusion
Revascularization, coronary artery [14]	RCT	1 (2687)	High	N/A	Imprecise	Direct	Single study	Insufficient	No conclusion
	adj NRCS	1 (295)	Moderate	N/A	Precise	Direct	Single study	Insufficient	No conclusion
	Overall	*2 (2982)*	*High*	*Inconsistent*	*Imprecise*	*Direct*	*None*	*Insufficient*	*No conclusion*
CHF[15]	RCT	2 (3410)	High / Moderate	N/A[16]	Highly imprecise	Direct	Single study[17]	Insufficient	No conclusion

[9] Evidence supporting the conclusion that CPAP may reduce the risk of death may be most applicable to older adults (at higher risk of death) and/or longer-term followup (up to 11 years).

[10] One study reported on risk of incident angina and one study reported on risk of hospitalization for unstable angina.

[11] One study reported on risk of incident angina and one study reported on risk of hospitalization for unstable angina.

[12] Study reporting on incident angina was highly imprecise. Study reporting on hospitalization for unstable angina was precise.

[13] One study reported on risk of incident angina and one study reported on risk of hospitalization for unstable angina.

[14] The evaluated outcome was repeat coronary artery revascularization among participants with a history of revascularization. A third study (RCT) also evaluated risk of any major artery revascularization; as a unique study, it provided insufficient evidence.

[15] One study reported on risk of incident CHF and one study reported on risk of hospitalization for CHF.

[16] One study reported on risk of incident CHF and one study reported on risk of hospitalization for CHF.

[17] One study reported on risk of incident CHF and one study reported on risk of hospitalization for CHF.

Outcome	Study Design	No. Studies (Subjects)	Risk of Bias	Consistency	Precision	Directness	Other	Overall SoE	Conclusion Statements Estimate (95% CI)
AFib	RCT	2 (2917)	Moderate	Inconsistent	Highly imprecise	Direct	Clin heterog	Insufficient	No conclusion
Composite CV outcomes	RCT	6 (4146)	High	Consistent	Imprecise	Indirect[18]	Clin heterog	Low	No evidence of effect ES range 0.42-1.10 (all NS)[19] **NNT range 10 to −63*)**
	adj NRCS	3 (2596)	Moderate	Variable[20]	Precise	Indirect[21]	Few studies	not evaluated	CPAP associated with lower risk (in patients not restricted by CV history) ~0.38[22] and 0.64 (sig)
	Overall	*9 (6742)*	*High*	*Inconsistent*	*Imprecise*	*Indirect[23]*	*Clin heterog*	*Low*	*No evidence of effect*

Abbreviations: adj NRCS = adjusted nonrandomized comparative studies, AFib = atrial fibrillation, AMI = acute myocardial infarction, Assn = association, CHF = congestive heart failure, CI = confidence interval, Clin heterog = clinically heterogeneous (e.g., different populations, eligibility criteria), CV = cardiovascular, ES = (summary) effect size, HR = (summary) hazard ratio, N/A = not applicable, 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.

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

* The negative value should be interpreted as a number-needed-to-harm (e.g., as an upper bound, for every 85 people treated with CPAP, one additional person would die).

[18] Each study with a unique (though overlapping) composite CVD outcome.

[19] Meta-analysis not conducted due to clinical heterogeneity of outcomes. This range excludes one highly imprecise estimate.

[20] One NRCS in patients with recent coronary revascularization found no significant association, nominally favoring no CPAP. Two NRCSs in patients regardless of CV history were consistent in finding a significant association favoring CPAP use.

[21] Each study with a unique (though overlapping) composite CVD outcome.

[22] Inverse of reported adjusted HR (which was analyzed as no CPAP vs. CPAP).

[23] Each study with a unique (though overlapping) composite CVD outcome.

Cardiovascular Mortality

Four studies (3 RCTs and 1 NRCS) comparing CPAP and no CPAP reported on incidence of CVD mortality after about 4 to 7 years (Table CVD.03).[13, 89, 94, 101, 102] 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 quasi-ITT) analysis (CPAP treatment/prescription vs. no treatment/prescription). One study also reported an as-treated analysis.

Randomized Controlled Trials (Cardiovascular Mortality)

The three 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).

In all RCTs, patients and clinicians were not blinded, but outcome assessors were blinded in Barbé 2012 and SAVE. The SAVE trial was, 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 AHI pre-treatment (Barbé 2012) or high crossover rate (RICCADSA).

Across trials, effect sizes ranged from 0.41 to 3.08, all statistically nonsignificant and imprecise. The NNT to prevent one CV death ranged from 31 to 357 across studies. The summary OR across the RCTs was highly imprecise at 0.99 (95% CI 0.43 to 2.25) (Figure CVD.02). Under an assumption of a control rate of 1.6 percent, the summary NNT was 6351 (95% CI 110 to −52). The SAVE trial reported both an ITT and a propensity score matched analysis of compliant CPAP users (sett Table CVD.03). The compliant analysis nominally favored CPAP, but was highly imprecise and statistically similar to the ITT analysis.

Adjusted Nonrandomized Comparative Study (Cardiovascular Mortality)

The single NRCS, Nakamura 2009, evaluated CPAP treated and untreated patients in a national registry.[101, 102] The study was deemed to of low risk of bias based on their use of appropriate propensity score matching between groups. However, as with other NRCSs, there may be unaccounted-for bias related to self-selection of CPAP use, despite the goal of replicating an ITT analysis. The NRCS reported a HR of 0.80 (95% CI 0.38 to 1.70) over a mean of 6.5 years of follow up (Figure CVD.02).

All Studies (Cardiovascular Mortality)

Combining this NRCS with the RCTs yielded a summary effect estimate of 0.97 (95% CI 0.62 to 1.53; I^2 =3%), suggesting no difference in CV mortality with or without CPAP treatment (Figure CVD.02). 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 analysis.

The three RCTs were statistically homogenous (I^2 =3%). 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 trial found no statistical difference in HR in its ITT and compliant users analyses. They also reported no significant association in *post hoc* CPAP dose-response analyses and CV end points. Across studies, there was no correlation between CPAP compliance and effect sizes.

Applicability (Cardiovascular Mortality)

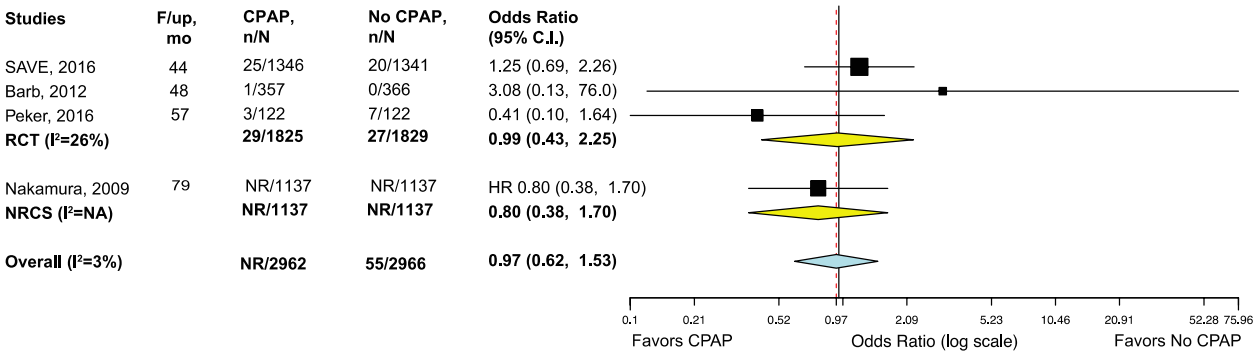
Among the RCTs, the majority of included participants (74%) were in the SAVE trial, which was restricted to adults with severe OSA (ODI ≥12) and preexisting CAD or CeVD. The effect estimates from the other two RCTs were highly imprecise. Thus, the summary estimate is most applicable to patients who would 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

Three RCTs provide insufficient evidence to determine the effect of CPAP on CV mortality (Table CVD.02). Study estimates, individually and by meta-analysis are highly imprecise (summary effect size 0.99 (95% CI 0.43 to 2.25).

However, combining the RCTs with the propensity-score matched NRCS provides low SoE that CPAP does not affect CV mortality (OR = 0.97, 95% CI 0.62 to 1.53). 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. It would not be unexpected for future evidence to alter these conclusions.

Figure CVD.02. Meta-analysis of CPAP versus No CPAP: Cardiovascular mortality



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%, NA = not applicable, NR = not reported, NRCS = nonrandomized comparative study, RCT = randomized controlled trial, SAVE = Sleep Apnea cardioVascular Endpoints trial.

Table CVD.03. CPAP versus no CPAP: Cardiovascular mortality

Study PMID	Design	Analysis (Powered?[24])	Followup Duration (mo)	Arm	n/N (%)	CPAP Compliance	Effect Size (95% CI) Calculated NNT (95% CI)[25]
Barbé 2012 22618923[89]	RCT	ITT	48	CPAP	1/357 (0.3)	64% overall	OR 3.08 (0.13, 76.0)*
		(No)		No CPAP	0/366 (0)		NNH 357 (NNT 373, 121)
RICCADSA 26914592[94]	RCT	ITT	57	CPAP	3/122 (2.5)	60% at 1 yr	OR 0.41 (0.10, 1.64)*
		(No)		No CPAP	7/122 (5.7)		NNT 31 (12, NNH 60)
SAVE 27571048[13]	RCT	ITT	44	CPAP	25/1346 (1.9)	42% overall	HR 1.22 (0.68, 1.64) NNH 273 (NNT 166, 75) [Compliant analysis: HR 0.90 (0.41, 2.01)]
		(No)		No CPAP	20/1341 (1.5)		
Nakamura 2009 No PMID[101, 102]	NRCS	Quasi-ITT	79	CPAP	NR/1137	NR	Adjusted HR 0.80 (0.38, 1.70)[26]
		(No)		No CPAP	NR/1137		

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.

* Calculated

[24] Explicitly powered for this outcome?

[25] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH (or NNT) in the 95% CI range suggests that one CV death may occur (or be prevented) per the reported number of people treated (within the 95% CI).

[26] Propensity score analysis: age, sex, lifestyle (smoking, drinking), comorbid conditions including cardiovascular disease and/or stroke, other comorbid conditions, ESS and AHI.

Stroke and Transient Ischemic Attack

Four RCTs comparing CPAP and no CPAP reported on incidence of stroke and/or TIA after 1 or about 4 years (Table CVD.04).[13, 89, 93, 94] 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 compliant CPAP users analyses.

The four RCTs evaluated different, but overlapping, sets of patients. Two trials (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. Three 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).

None of the studies blinded patients or their clinicians, but the SAVE and PREDICT trials were deemed to be at moderate risk of bias since outcome assessors were blinded and there were no other important biases evident. Barbé 2012 and RICCADSA were downgraded to high risk of bias for failure to adjust for baseline differences in AHI (Barbé 2012) or high crossover rate (RICCADSA).

Stroke

All four trials evaluated risk of stroke (Figure CVD.03). Effect sizes ranged from 0.49 to 1.54 across studies, all statistically nonsignificant. NNT ranged from 41 to a NNH of 340 across studies. Most trials were highly imprecise. The summary OR was 0.96 (95% CI 0.69 to 1.33; 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 2.2 percent, the summary NNT was 1161 (95% CI 149 to −142). In a propensity score matched analysis of compliant CPAP users, the SAVE trial reported a stronger effect size that was just statistically significant (see Table CVD.04), but the study did not report whether this analysis was significantly different than its primary ITT analysis.

Transient Ischemic Attack

Three trials evaluated risk of TIA (Figure CVD.04). Among the three trials, SAVE specified the outcome as hospitalization due to TIA. Two of the trials were highly imprecise; the SAVE trial found somewhat more patients in the CPAP group were hospitalized for TIA than in the no CPAP group. Effect sizes ranged from 0.41 to 1.78 and NNT ranged from 120 to a NNH of 193 across trials. Across all three trials (combining TIA and hospitalization for TIA) the summary effect size was 0.95 (95% CI 0.31 to 2.87; I^2 =40%), which was highly imprecise, precluding a conclusion regarding the effect of CPAP on risk of TIA. Under an assumption of a control rate of 1.1 percent, the summary NNT was 1837 (95% CI 132 to −50). The SAVE trial's analysis of compliant CPAP users yielded an even more imprecise estimate (Table CVD.04)

Heterogeneity of Treatment Effect (Stroke and TIA)

None of the studies reported subgroup analyses. For stroke, the trials were mostly highly imprecise and homogeneous, no potential explanation of heterogeneity could be elucidated across studies. While there was some heterogeneity in TIA findings across studies, again no potential explanation of heterogeneity could be elucidated. As noted, the SAVE trial found a

stronger effect on stroke in its analyses of compliant 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)

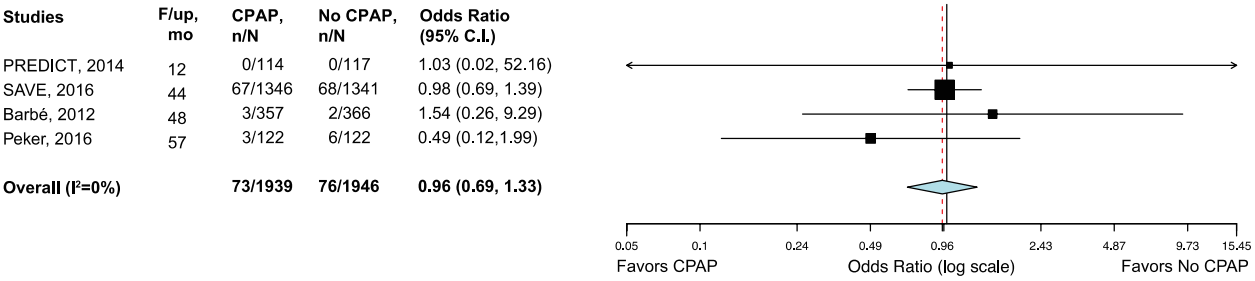
The findings are likely most applicable to patients who would meet eligibility criteria for the SAVE trial, which was restricted to adults with severe OSA (ODI ≥12) and preexisting CAD or CeVD. The effect estimates from the other two RCTs were highly imprecise. Although, their findings were imprecise, 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

Four RCTs provide low SoE that CPAP does not affect the risk of stroke (Table CVD.02), based on a somewhat imprecise estimate of effect (summary OR 0.96, 95% CI 0.69 to 1.33). 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. It would not be unexpected for future evidence to alter these conclusions.

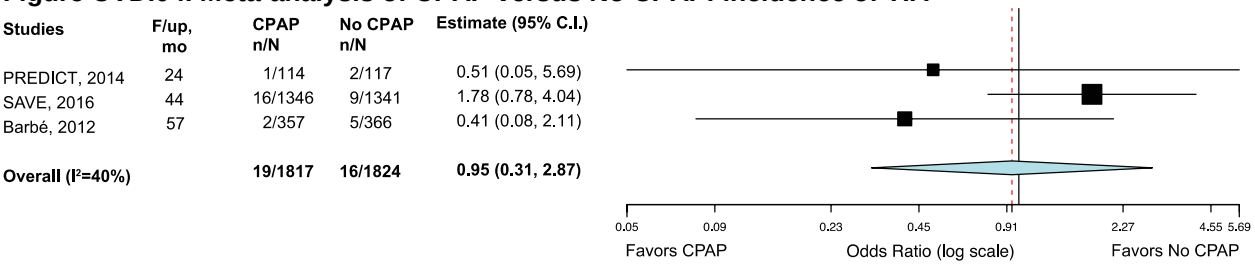
Three RCTs yielded a highly imprecise estimate of the effect of CPAP on risk of TIA and, thus, provide insufficient evidence.

Figure CVD.03. Meta-analysis of CPAP versus No CPAP: incidence of stroke



Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, I^2 = measure of statistical heterogeneity ranging from 0% (none) to 100%, PREDICT = trial with undefined acronym, RCT = randomized controlled trial, SAVE = Sleep Apnea cardioVascular Endpoints trial.

Figure CVD.04. Meta-analysis of CPAP versus No CPAP: incidence of TIA



Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, I^2 = measure of statistical heterogeneity ranging from 0% (none) to 100%, PREDICT = trial with undefined acronym, RCT = randomized controlled trial, Sleep Apnea cardioVascular Endpoints trial, TIA = transient ischemic attack.

Table CVD.04. CPAP versus no CPAP: Stroke and TIA

Study PMID	Design	Outcome	Analysis (Powered?[27])	Followup Duration (mo)	Arm	CPAP Compliance	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI)[28]
Barbé 2012 22618923[89]	RCT	Stroke	ITT (No)	48	CPAP	64% overall	3/357 (0.8)	OR 1.54 (0.26, 9.29)*
					No CPAP		2/366 (0.5)	NNH 340 (NNT 109, 66)
RICCADSA 26914592[94]	RCT	Stroke	ITT (No)	57	CPAP	60% at 1 yr	3/122 (2.5)	OR 0.49 (0.12, 2.00)*
					No CPAP		6/122 (4.9)	NNT 41 (14, NNH 44)
PREDICT 25172769[93]	RCT	Stroke	ITT (No)	12	CPAP	NR	0/114 (0)	OR 1.03 (0.02, 52.2)*
					No CPAP		0/117 (0)	Not calculable
SAVE 27571048[13]	RCT	Stroke	ITT (No)	44	CPAP	42% overall	67/1346 (5.0)	HR 0.97 (0.69, 1.35) NNT 1074 (57, NNH 64)
					No CPAP		68/1341 (5.1)	[Compliant analysis: HR 0.56 (0.32, 1.00)]
Barbé 2012 22618923[89]	RCT	TIA	ITT (No)	48	CPAP	64% overall	2/357 (0.6)	OR 0.41 (0.08, 2.11)*
					No CPAP		5/366 (1.4)	NNT 124 (45, NNH 163)
PREDICT 25172769[93]	RCT	TIA	ITT (No)	12	CPAP	NR	1/114 (0.9)	OR 0.51 (0.05, 5.69)*
					No CPAP		2/117 (1.7)	NNT 120 (27, NNH 48)
SAVE 27571048[13]	RCT	Hospitalization for TIA	ITT (No)	44	CPAP	42% overall	16/1346 (1.2)	HR 1.88 (0.83, 4.28) NNH 193 (NNT 481, 80)
					No CPAP		9/1341 (0.7)	[Compliant 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.

* Calculated

[27] Explicitly powered for this outcome?

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

Acute Myocardial Infarction

Four RCTs comparing CPAP and no CPAP reported on incidence of acute MI after 1 or about 4 years (Table CVD.05). [13, 89, 93, 94] 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 reported a propensity score matched compliant CPAP users analysis.

The four RCTs evaluated different, but overlapping, sets of patients. Two trials (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. Three 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).

None of the studies blinded patients or their clinicians, but the SAVE and PREDICT trials were deemed to be at moderate risk of bias since outcome assessors were blinded and there were no other important biases evident. Barbé 2012 and RICCADSA were downgraded to high risk of bias for failure to adjust for baseline differences in AHI (Barbé 2012) or high crossover rate (RICCADSA).

Across the four RCTs there was a wide range of effect sizes estimates from 0.25 to 7.38, but estimates were all 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 1.06 (95% CI 0.72 to 1.56) (Figure CVD.05). 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 2.5 percent, the summary NNT was −685 (95% CI 145 to −74); i.e., a NNH of 685.

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 compliant CPAP users than its ITT analysis (Table CVD.05). 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 trial, which was restricted to adults with severe OSA (ODI ≥12) and preexisting CAD or CeVD. The effect estimates from the other two 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.

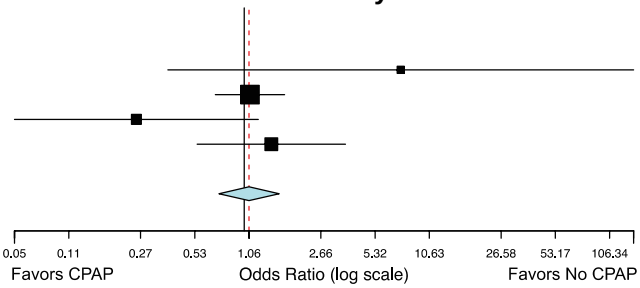
Summary of Effect of CPAP on Acute Myocardial Infarction

Four trials provide low SoE that CPAP does not affect risk of acute MI (Table CVD.02). 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 OR 1.06, 95% CI 0.72 to 1.56). 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. It would not be unexpected for future evidence to alter these conclusions.

Figure CVD.05. Meta-analysis of CPAP versus No CPAP: Incidence of acute myocardial infarction

Studies	F/up, mo	CPAP, n/N	No CPAP, n/N	Odds Ratio (95% C.I.)
PREDICT, 2014	12	3/114	0/117	7.38 (0.38, 144.43)
SAVE, 2016	44	42/1346	39/1341	1.08 (0.69, 1.67)
Barbé, 2012	48	2/357	8/366	0.25 (0.05, 1.20)
Peker, 2016	57	11/122	8/122	1.41 (0.55, 3.64)
Overall (I^2=0%)		**58/1939**	**55/1946**	**1.06 (0.72, 1.56)**



Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, PREDICT = trial with undefined acronym, Sleep Apnea cardioVascular Endpoints trial.

Table CVD.05. CPAP versus no CPAP: Acute myocardial infarction

Study PMID	Design	Analysis (Powered?[29])	Followup Duration (mo)	Arm	CPAP Compliance	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI)[30]
Barbé 2012 22618923[89]	RCT	ITT	48	CPAP	64% overall	2/357 (0.6)	OR 0.25 (0.05, 1.20)*
		(No)		No CPAP		8/366 (2.2)	NNT 62 (30, NNH 1647)
PREDICT 25172769[93]	RCT	ITT	12	CPAP	NR	3/114 (2.6)	OR 7.38 (0.38, 144)*
		(No)		No CPAP		0/117 (0)	NNH 38 (NNT 326, 18)
RICCADSA 26914592[94]	RCT	ITT	57	CPAP	60% at 1 yr	11/122 (9.0)	OR 1.41 (0.55, 3.64)*
		(No)		No CPAP		8/122 (6.6)	NNH 41 (NNT 23, 11)
SAVE 27571048[13]	RCT	ITT	44	CPAP	42% overall	42/1346 (3.1)	HR 1.06 (0.68, 1.64) NNH 472 (NNT 93, 66)
		(No)		No CPAP		39/1341 (2.9)	[Compliant 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.

* Calculated

[29] Explicitly powered for this outcome?

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

Angina

Two RCTS reported on risk of angina. The PREDICT trial reported on incidence of angina at 1 year[93] and the SAVE trial reported on incidence of hospitalization for unstable angina at 3.7 years (Table CVD.06).[13] The PREDICT trial was restricted to older adults (≥65 years), regardless of CVD history while the SAVE trial included only adults with existing CVD or CeVD. The two trials were deemed to be at moderate risk of bias. Neither trial was explicitly powered for angina. Both reported ITT analyses; SAVE also reported a propensity score matched compliers 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).[93] SAVE found no significant difference in risk of hospitalization for unstable angina (HR 1.09; 95% CI 0.82 to 1.45; NNH 155).[13] The analysis CPAP compliers yielded a similar estimate (Table CVD.06). SAVE also reported no significant association in *post hoc* CPAP dose-response analyses and CV end points.

Heterogeneity of Treatment Effect (Angina)

Neither of the studies reported subgroup analyses. Participants in the SAVE trial (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 between studies, given the high degree of imprecision in the PREDICT trial. As noted, the effect of CPAP did not vary based on compliance.

Applicability (Angina)

The findings are most applicable to patients who would meet eligibility criteria for the SAVE trial, which was restricted to adults with severe OSA (ODI ≥12) and preexisting CAD or CeVD. The effect estimates from the other RCT (of older adults).

Summary of Effect of CPAP on Angina

With only a single RCT evaluating each outcome, there is insufficient evidence to determine the effect of CPAP on risk of incident angina or hospitalization for unstable angina (Table CVD.02).

Table CVD.06. CPAP versus no CPAP: Angina

Study PMID	Design	Outcome	Analysis (Powered ?[31])	Followup Duration (mo)	Arm	CPAP Compliance	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI)[32]
PREDICT 25172769[93]	RCT	Angina	ITT (No)	12	CPAP	NR	2/113 (1.8)	OR 0.68 (0.11, 4.18)* NNT 126 (22, NNH 34)
					No CPAP		3/117 (2.6)	
SAVE 27571048[13]	RCT	Hospitalization for unstable angina	ITT (No)	44	CPAP	42% overall	99/1346 (7.4)	HR 1.09 (0.82, 1.45) NNH 155 (NNT 78, 39) [Compliant analysis: HR 0.99 (0.64, 1.51)]
					No CPAP		90/1341 (6.7)	

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.

* Calculated

Revascularization

Three studies (two RCTs and one NRCS) comparing CPAP and no CPAP reported on incidence of revascularization (Table CVD.07).[13, 94, 105] The RICCADSA trial and the Wu 2015 NRCS included only patients with prior coronary artery revascularization. These studies evaluated the risk of repeat coronary artery revascularization. The SAVE RCT, in contrast, 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 3.7 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 two RCTs evaluated different outcomes (repeat coronary artery revascularization and any major arterial revascularization). Neither of the RCTs blinded patients or their clinicians, but SAVE blinded outcome assessors and was deemed to be at moderate risk of bias. RICCADSA was deemed to be at high risk of bias due to a high crossover rate. Neither trial was explicitly powered for a revascularization outcome. Both reported ITT analyses, but SAVE also reported a *post hoc* propensity score matched analysis of CPAP compliers.

The SAVE trial found no significant difference in risk of any arterial revascularization 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 CVD.07).[13] The CPAP compliers analysis found a similar result.

[31] Explicitly powered for this outcome?

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

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.[94]

Adjusted Nonrandomized Comparative Study (Revascularization)

The Wu 2015 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 for this outcome; the trial compared compliant CPAP users with those who refused CPAP or were noncompliant and was, thus, an "as-treated" analysis. 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.[105] Particularly as an as-treated analysis, this NRCS may have been subject to uncontrolled biases related to self-selection of CPAP use and compliance. While we did not formally combine the RCTs and NRCS, it can be noted that in our sensitivity analysis of progressively borrowing information from NRCSs to the summary estimate of RCTs, the summary estimate between RICCADSA and Wu 2015 shifts from nominally favoring no CPAP (as in the ITT RCTs) to nominally favoring CPAP (consistent with the as-treated NRCS); however, the estimates are mostly highly imprecise.

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 who would be eligible for enrollment in the SAVE trial, i.e., adults with severe OSA (ODI ≥12) and preexisting CAD or CeVD. By definition, the findings regarding the risk of repeat coronary artery revascularization are 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

With only a single study evaluating any (major) revascularization (coronary, peripheral, or carotid arteries), there is insufficient evidence to make a conclusion regarding the effect of CPAP on these outcomes (Table CVD.02).

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 NRCS shows a significantly higher adjusted HR for revascularization in the no CPAP group.

Table CVD.07. CPAP versus no CPAP: Revascularization

Study PMID	Design	Outcome	Analysis (Powered?[33])	Followup Duration (mo)	Arm	CPAP Compliance	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI)[34]
RICCADSA 26914592[94]	RCT	Repeat coronary revascularization	ITT	57	CPAP	60% at 1 yr	17/122 (13.9)	OR 1.25 (0.59, 2.66)*
			(No)		No CPAP		14/122 (11.5)	NNH 41 (NNT 17, 9)
SAVE 27571048[13]	RCT	Any revascularization	ITT	44	CPAP	42% overall	69*/1346 (5.1)	OR 1.31 (0.91, 1.89)* NNH 85 (NNT 250, 36)
			(No)		No CPAP		53*/1341 (4.0)	[Compliant analysis: HR 1.25 (0.79, 1.96)]
Wu 2015 25412159[105]	NRCS	Repeat coronary revascularization	As-treated	60	CPAP	100% (as-treated)	15/128 (11.7)	***No CPAP vs. CPAP*: Adjusted HR 2.13 (1.19, 3.81)[35]**
			(No)		No CPAP		35/167 (21.0)	

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.

* Calculated

[33] Explicitly powered for this outcome?

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

[35] Note that 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, PCI type (i.e., emergency vs elective),adjunctive medical therapy.

Congestive Heart Failure

Two RCTs comparing CPAP and no CPAP reported on incidence of either congestive heart failure (CHF)[89] or hospitalization for CHF (Table CVD.08).[13] The two RCTs evaluated distinct populations: Barbé 2012 included relatively young (≤70 years, mean age 52), otherwise healthy participants; SAVE included patients with known histories of either CVD or CeVD who were, on average, older (mean age 61 years). Nevertheless, incident CHF was similar and relatively uncommon in the two RCTs (about 1%).

In both RCTs, patients and clinicians were not blinded. The SAVE trial was deemed to be at moderate risk of bias since outcome assessors were blinded. 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. Neither trial was explicitly powered for a CHF outcome. Both reported ITT analyses; SAVE also reported a *post hoc* propensity score matched analysis of CPAP compliers.

Barbé 2012 reported the incidence of CHF at a mean Followup of 4 years. Only 1.1 percent of patients developed CHF and the difference between treatment groups was highly imprecise (OR 0.61, 95% CI 0.15 to 2.58; NNT 190).[89]

The SAVE trial reported the incidence of hospitalization for CHF at a mean Followup of 3.7 years. CHF hospitalization was also uncommon in this RCT (1.3%), with an almost highly imprecise estimate of the difference between treatment groups (HR 0.98, 95% CI 0.50 to 1.92; NNT 21,235).[13] The analysis of CPAP compliers yielded very similar results (Table CVD.08).

Heterogeneity of Treatment Effect (CHF)

Neither RCT reported subgroup analyses. SAVE reported no significant association in *post hoc* CPAP dose-response analyses and CV end points. The two RCTs were imprecise, given the relatively low incidence of CHF; thus, no differences could be discerned between trials.

Applicability (CHF)

The two 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 only a single RCT evaluating each outcome, there is insufficient evidence to determine the effect of CPAP on risk of CHF or hospitalization for CHF (Table CVD.02).

Table CVD.08. CPAP versus no CPAP: Congestive heart failure

Study PMID	Design	Outcome	Analysis (Powered?[36])	Followup Duration (mo)	Arm	CPAP Compliance	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI)[37]
Barbé 2012 22618923[89]	RCT	CHF	ITT (No)	48	CPAP	64% overall	3/357 (0.8)	OR 0.61 (0.15, 2.58)* NNT 190 (49, NNH 101)
					No CPAP		5/366 (1.4)	
SAVE 27571048[13]	RCT	Hospitalization for CHF	ITT (No)	44	CPAP	42% overall	17/1346 (1.3)	HR 0.98 (0.50, 1.92) NNT 21,235 (118, NNH 119) [Compliant analysis: HR 0.82 (0.34, 2.03)]
					No CPAP		17/1341 (1.3)	

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.

* Calculated

Atrial Fibrillation

Two RCTS reported on risk of AFib. The PREDICT trial reported on incidence of AFib at 1 year[93] and the SAVE trial reported 3.7 year followup (Table CVD.09).[13] The PREDICT trial was restricted to older adults (≥65 years), regardless of CVD history while the SAVE trial included only adults with existing CVD or CeVD. The two trials were deemed to be at moderate risk of bias. Both trials were conducted by ITT analysis, but SAVE also reported a propensity score matched analysis of CPAP compliers. Neither was explicitly powered for AFib.

The two RCTs both found imprecise estimates of effect, but with nominal inconsistency in direction of effect. In PREDICT, the OR was 0.48 (95% CI 0.19 to 1.24; NNT 17) at 1 year, [93] while in SAVE, the OR was 1.47 (95% CI 0.76 to 2.84; NNH 194).[13] The SAVE CPAP compliers analysis found a similar result. Pooling the two trials yielded a highly imprecise summary estimate (OR 0.89, 95% CI 0.30 to 2.63).

Heterogeneity of Treatment Effect (AFib)

Neither 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) had a higher incidence of AFib than the SAVE trial (of those at increased CV risk). It is unclear whether differences in the characteristics of the studies' participants explain any possible difference in effect of CPAP between trials; however, both trials found no significant effect of CPA on risk AFib.

[36] Explicitly powered for this outcome?

[37] 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).

Applicability (AFib)

The two 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 Atrial Fibrillation

Two RCTs provide insufficient evidence regarding the effect of CPAP on risk of AFib (Table CVD.02). Each trial was imprecise and the two trials yielded inconsistent estimates of effect.

Table CVD.09. CPAP versus no CPAP: Atrial fibrillation

Study PMID	Design	Analysis (Powered?[38])	Followup Duration (mo)	Arm	CPAP Compliance	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI)[39]
PREDICT 25172769[93]	RCT	ITT (No)	12	CPAP	NR	7/114 (6.1)	OR 0.48 (0.19, 1.24)* NNT 17 (8, NNH 66)
				No CPAP		14/117 (12.0)	
SAVE 27571048[13]	RCT	ITT (No)	44	CPAP	42% overall	22/1346 (1.6)	OR 1.47 (0.76, 2.84)* NNH 194 (NNT 274, 72) [Compliant analysis: HR 1.84 (0.74, 4.55)]
				No CPAP		15/1341 (1.1)	

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.

* Calculated

Composite Cardiovascular Outcomes

Nine studies (6 RCTs and 3 NRCSs) comparing CPAP and no CPAP reported on various composite CV outcomes (Tables CVD.10, CVD.11).[13, 89, 92-94, 105, 110, 113, 116] Studies evaluated a wide range of followup times (from 2 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)

Six RCTs evaluated eight unique (though overlapping) composite CV outcomes, as described in Table CVD.10, 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).

Three 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, 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).

[38] Explicitly powered for this outcome?

[39] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH (or NNT) in the 95% CI range suggests that one CV death may occur (or be prevented) per the reported number of people treated (within the 95% CI).

Two of the RCTs based their power calculations on composite CV outcomes (SAVE, RICCADSA); however, RICCADSA was ultimately underpowered. 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 noncompliant 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 noncompliant 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, heart failure, or TIA. However, due in part to recruitment difficulties and a blinded interim analysis that found better-than-expected CPAP compliance 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.

Five of the six 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 compliance (mean CPAP use <4 hours/night). Among the studies reporting ITT analyses, between about 38 and 64 percent of those assigned to CPAP were compliant (≥4 hours per night). Three of the ITT trials also reported analyses among compliant 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 CVD.10). 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; NNT 9). The three ITT RCTs that also reported sensitivity analyses among just CPAP compliers (SAVE, Barbé 2012, RICCADSA) all found nominally stronger effect sizes in the CPAP complier analysis, but the difference between analyses varied across studies (Table CVD.10). In SAVE, CPAP-complier based analyses of the primary outcome and composite ischemic CV events shifted the nominal 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 compliers analyses; their comparison of rate ratios among compliers and among noncompliers 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 (compliant vs. noncompliant and nonusers) analysis (HR 0.29, 95% CI 0.10 to 0.86).

A rough comparison between ITT and compliant/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). 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 (note that Table CVD.10 is ordered roughly by CV risk based on history of CVD and age). There was no correlation between compliance rate and effect size.

Adjusted Nonrandomized Comparative Studies (Composite Cardiovascular Outcomes)

Three NRCSs evaluated four distinct (but overlapping) composite CV outcomes (also distinct from the RCTs). One NRCS (Wu 2015) evaluated a population with recent coronary revascularization (Table CVD.11), while the other two NRCSs evaluated otherwise healthy participants (other than OSA). Outcomes were assessed at approximately 5 to 6 years. Two studies (Wu 2015, Schipper 2017) were deemed to be at moderate risk of bias for failure to use propensity score matching (or equivalent methodology); the third NRCS (Myllylä 2019) also used a regression analysis, but provided an incomplete description of participants' criteria for OSA or CPAP use and of the CPAP intervention; thus, the study was deemed to be at high risk of bias.

Like all other NRCSs, none was explicitly powered for any outcome, including the composite CV outcome. All three were reported "as-treated" analyses, comparing compliant CPAP users with a combination of nonusers and noncompliant (<4 hours/night) participants. Particularly as as-treated analyses, these NRCSs may have been subject to uncontrolled biases related to self-selection of CPAP use and compliance.

The three NRCSs all found adjusted hazard ratios that nominally favored CPAP use. 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 one NRCS conducted in patients with preexisting CVD (Wu 2015) was 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. Comparisons of CPAP use versus nonuse ("as-treated") mostly found marginally stronger associations among the as-treated studies (ES 0.18 to 0.83) than the ITT analyses (ES 0.42 to 1.10), but the difference in effect sizes by did not differ between sets of studies (P = 0.19). SAVE reported no significant association in *post hoc* CPAP dose-response analyses and CV end points.

Applicability (Composite Cardiovascular Outcomes)

The findings are generally applicable to a range 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

Six RCTs provide low SoE that CPAP does not affect the risk of composite CV outcomes (Table CVD.02). 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/noncompliance) and reduced risk of (variable) composite CV outcomes. A third NRCS, conducted in patients with recent coronary revascularization, 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 as-treated analyses, the NRCS may have been subject to uncontrolled biases related to self-selection of CPAP use and compliance.

Overall, we concluded that studies have not provided evidence that CPAP affects the risk of composite CV outcomes. Thus, overall, there is low SoE that CPAP does not affect the risk of composite CV outcomes.

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. It would not be unexpected for future evidence to alter these conclusions.

Table CVD.10. CPAP versus no CPAP: Randomized controlled trials reporting composite cardiovascular outcomes

Study PMID	Outcome	Population[40]	Analysis (Powered?[41])	Followup Duration CPAP Compliance	CPAP n/N (%)	No CPAP n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI)[42]
Huang 2015 25125635[92]	CeV death, cor revasc, CV death, HF hosp, MI, stroke	CVD 45-75 yo	As-treated (No)	Median 3 yr 100% (as treated)	1/36 (2.8)	5/37 (13.5)	OR 0.18 (0.02, 1.65)[43] NNT 9 (4, NNH 66)
SAVE 27571048[13]	CV death, HF hosp, MI, stroke hosp, TIA, UA[44]	CVD 45-75 yo	ITT (Yes)	Mean 3.7 yr 42% overall	229/1346 (17.0)	207/1341 (15.4)	HR 1.10 (0.91, 1.32) NNH 63 (NNT 83, 23) [Compliers analysis: HR 0.80 (0.60, 1.07)]
	CV death, MI, stroke				117/1346 (8.7)	120/1341 (8.9)	HR 0.96 (0.74, 1.23) NNT 390 (42, NNH 53) [HR 0.69 (0.46, 1.04)]
	Angina hosp, CV death, MI, stroke (ischemic), TIA hosp				207/1346 (15.4)	191/1341 (14.2)	HR 1.07 (0.88, 1.31) NNH 88 (NNT 65, 26) [HR 0.81 (0.59, 1.10)]
RICCADSA, 26914592[94]	Cor revasc, CV death, MI, stroke death[45]	CVD[46] Any age	ITT (Under-powered)	Mean 4.7 yr 60% at 1 yr	22/122 (18.1)	27/122 (22.1)	HR 0.62 (0.34, 1.13) NNT 24 (7, NNH 17) [**As-treated: HR 0.29 (0.10, 0.86**)]
MOSAIC 23111478[116]	AFib, angina, cor revasc, CV death, DM, HTN, MI, PVD, stroke, TIA	General 45-75 yo	ITT (No)	2 yr 38% at 6 mo	8/94 (8.5)	17/94 (18.1)	**OR 0.42 (0.18, 0.996)**[47] NNT 10 (5, NNH 2763)
PREDICT 25172769[93]	AFib, angina, MI, PVD, stroke, TIA	General ≥65 yo	ITT (No)	1 yr NR	14/114 (12.3)	17/117 (14.5)	HR 0.87 (0.40, 1.88) NNT 44 (9, NNH 15)
Barbé 2012 22618923[89]	Arrhythmia, CV death, HF, MI, stroke, TIA, UA	No CVD ≤70 yo	ITT (No)	4 yr 64% overall	28/357 (7.8)	31/366 (8.5)	Rate ratio 0.87 (0.52, 1.45) NNT 160 (22, NNH 30) [Compliers analysis: HR 0.80 (0.45, 1.43)]

[40] Based on study eligibility criteria. General = general population, no restriction related to CVD.

[41] Explicitly powered for this outcome?

[42] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH (or NNT) in the 95% CI range suggests that one CV death may occur (or be prevented) per the reported number of people treated (within the 95% CI).

[43] Calculated

[44] Power calculation for study based on this composite outcome.

[45] Power calculation for study based on this composite outcome.

[46] Coronary revascularization

[47] Calculated

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).

Table CVD.11. CPAP versus no CPAP: Adjusted nonrandomized comparative studies reporting composite cardiovascular outcomes

Study PMID	Outcome	Population [48]	Analysis (Powered ?[49])	Followup Duration	CPAP n/N (%)	No CPAP n/N (%)	Effect Size (95% CI)
Wu 2015 25412159[105]	Cor revasc, death, MI, stent thrombosis, stroke	CVD[50] Any age	As-treated	Median 4.8 yr	40/128 (31.3)	59/167 (35.3)	Adj HR *(no CPAP vs. CPAP)* 1.22 (0.79, 1.87)[51]
	Cor revasc, death, MI, stent thrombosis		(No)		30/128 (23.4)	52/167 (31.1)	Adj HR *(no CPAP vs. CPAP)* 1.52 (0.94, 2.56)[52]
Myllylä 2019 30848437[113]	Angina, CAD (incident), CAD death, CM death, MI, stroke, stroke death, other CV/vascular death[53]	General Any age	As-treated (No)	Median 6.2 yr	148/1030 (14.4)	194/1030 (18.84)	**Adj HR 0.64 (0.5, 0.8)**
Schipper 2017 28550476[110]	MI, stroke, TIA	General Any age	As-treated (No)	5.9 yr	6/140 (4.3)	11/101 (10.9)	**Adj HR *(no CPAP vs. CPAP)* 2.66 (1.20, 5.91)[54]**

Statistically significant results are in bold font.

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

[48] Based on study eligibility criteria. General = general population, no restriction related to CVD.

[49] Explicitly powered for this outcome?

[50] Coronary revascularization

[51] Note that this HR is for *no CPAP versus CPAP*, opposite of typical analysis direction. The calculated (unadjusted) OR for CPAP vs. no CPAP was 0.83 (0.51, 1.36).

[52] Note that this HR is for *no CPAP versus CPAP*, opposite of typical analysis direction. The calculated (unadjusted) OR for CPAP vs. no CPAP was 0.68 (0.40, 1.14).

[53] Death due to aortic dissection, aortic aneurysm, arrhythmia, heart valve disease, atherosclerosis, hypertension, or deep vein thrombophlebitis.

[54] Note that this HR is for *no CPAP versus CPAP*, typical analysis direction. The calculated (unadjusted) OR for CPAP vs. no CPAP was 0.37 (95% CI 0.13, 1.03).

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 and clinician bias; although, outcome assessors were blinded. Neither trial was explicitly powered for accident-related outcomes. Both were evaluated as ITT analyses. The SAVE trial reported about 42 percent compliance; PREDICT did not report on compliance.

Across specific outcome measures (Table X.01), the PREDICT trial,[93] at 1 year followup, and the SAVE trial,[13] at 3.7 years followup, found no statistically significant differences between groups using CPAP or not; although one measure (the annual rate of accident-causing injuries) in the SAVE trial was near significant, 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 compliance and accident rates.

Summary of Effect of CPAP on Accidents

Two RCTs provide low SoE of no effect of CPAP on risk of traffic or driving accidents (Table X.02). 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. It would not be unexpected for future evidence to alter these conclusions.

The evidence was insufficient regarding other types of accidents.

Table X.01. CPAP versus no CPAP: Accidents

Study PMID	Outcome	Analysis (Powered?[55])	Followup Duration (mo)	Arm	Baseline Rate	Final Rate	Effect Size (95% CI) Calculated NNT (95% CI)[56]
SAVE 27571048[13] 27571048	Traffic accidents, participants	ITT	44.4 (mean)	CPAP	NR	41/1346 (3.0%)	OR 0.86 (0.57, 1.32) [57]
		(No)		No CPAP	NR	47/1341 (3.5%)	NNT 218 (55, NNH 113)
	Traffic accidents, annual rate	ITT	44.4 (mean)	CPAP	NR	56 (1.1%/yr)	Rate ratio 0.78 (0.55, 1.11)
		(No)		No CPAP	NR	70 (1.4%/yr)	NNT 325/yr (128, NNH 649)
	Accident-causing injury, participants	ITT	44.4 (mean)	CPAP	NR	99/1346 (7.4%)	OR 0.82 (0.62, 1.09) [58]
		(No)		No CPAP	NR	118/1341 (8.8%)	NNT 69 (29, NNH 162)
	Accident-causing injury, annual rate	ITT	44.4 (mean)	CPAP	NR	219 (4.4%/yr)	Rate ratio 0.84 (0.70,1.00)[59]
		(No)		No CPAP	NR	255 (5.2%/yr)	NNT 120/yr (64, NC)
PREDICT 25172769[93]	Driving accidents	ITT	12	CPAP	1/73 (1.4)	2/73 (2.7)	OR 0.52 (0.05, 5.87)[60]
		(No)		No CPAP	2/77 (2.6)[1]	1/77 (1.3)	NNT 81 (18, NNH 31)
	Home accidents[61]	ITT	12	CPAP	12/117 (6.7)[1]	9/113 (8.0)	"Treatment effect" 0.49 (0.21,1.18)
		(No)		No CPAP	6/113 (5.3)[1]	18/117 (15.4)	NNT 13 (6, NNH 124)

Abbreviations: CI = confidence interval, CPAP = continuous positive airway pressure (device), ITT = intention-to-treat, NC = not calculable, NNH = number needed to harm, NNT = number needed to treat, OR = odds ratio, PREDICT = trial with undefined acronym, SAVE = Sleep Apnea cardioVascular Endpoints trial.

[55] Explicitly powered for this outcome?

[56] 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).

[57] Calculated

[58] Calculated

[59] P=0.06

[60] Calculated, 12 month data only

[61] The proportion of patients experiencing any accidents was analyzed adjusting accident history at baseline.

Table X.02. 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[62]
Hypertension	RCT	2 (423)[63]	High/Moderate[64]	N/A	Precise	Direct	Sparse[65]	Insufficient	No conclusion
Diabetes	RCT	2 (2733)	Moderate	Consistent	Imprecise	Direct	Clin heterog	Low	No evidence of an effect ES 0.85 (0.61, 1.18) **NNT 74 (28, −63**[66])
	adj NRCS	1 (266)	High	N/A	Precise	Indirect[67]	Sparse	not evaluated	No conclusion
	Overall	*3 (2999)*	*Moderate*	*Consistent*	Precise	*Direct*	*Clin heterog*	*Low*	*No evidence of an effect ES 0.75 (0.51, 1.08)*
Depression	RCT	4 (2824)	Moderate	Consistent	Precise	Indirect	Clin heterog	Low	No clinically significant improvement in depression symptom score
Anxiety	RCT	3 (2678)	Moderate	Consistent	Precise	Indirect	Clin heterog	Low	No clinically significant improvements in anxiety scores

[62] 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.

[63] One trial each reported on incident HTN and resolution of HTN.

[64] One trial each reported on incident HTN and resolution of HTN.

[65] One trial each reported on incident HTN and resolution of HTN.

[66] The negative value should be interpreted as a number-needed-to-harm (e.g., as an upper bound, for every 63 people treated with CPAP, one additional person develop diabetes).

[67] Evaluation of "regular" CPAP use vs. not regular use.

Outcome	Study Design	No. Studies (Subjects)	Risk of Bias	Consistency	Precision	Directness	Other	Overall SoE	Conclusion Statements Estimate (95% CI)
Cognitive function	RCT	3 (457)[68]	Moderate	Inconsistent[69]	Imprecise	Indirect	Clin heterog	Low	No evidence of a clinically significant effect on MMSE or TMT-B[70]
	adj NRCS	1 (126)	Moderate	N/A	Precise	Indirect	Sparse	not evaluated	No conclusion
	Overall	*4 (583)*	*Moderate*	*Inconsistent[71]*	*Imprecise*	*Indirect*	*Clin heterog*	*Low*	*Same as RCT conclusions*
QoL & functional status	RCT	8 (4107)[72]	Moderate-High	PCS: Inconsistent Others: Consistent	Precise	Indirect	None	Low	PCS: No effect. MCS, EuroQol-5D, and SAQLI: Small, not clinically significant improvements[73]
	Adj NRCS	1 (214)	High	N/A	Precise	Indirect	Sparse	not evaluated	No conclusion
	Overall	*9 (4321)[74]*	*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, 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.

[68] A fourth RCT (N=1098) evaluated unique measures of cognitive function and thus, provided only insufficient evidence.

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

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

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

[72] Across all QoL and functional status measures

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

[74] Across all QoL and functional status measures

Hypertension

Two RCTs compared the effect of CPAP versus no CPAP on incident HTN[89] or reversion to normotension.[92] Both Barbé 2012 and Huang 2015 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.

Barbé 2012 reported both an ITT and a CPAP compliers analysis, with 59 percent of analyzed participants compliant 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). Huang 2015 conducted an "as-treated" comparison of compliant 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).

Incident Hypertension

Barbé 2012 reported on the incidence of HTN after a 4 year followup period (Table X.03). 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 compliant CPAP users yielded a similar result (see Table X.03). 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 X.03). 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 compliant 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.

Heterogeneity of Treatment Effect (HTN)

Neither study reported subgroup analyses for either 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 is most applicable to patients with uncontrolled HTN and CAD who were compliant 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 or reversion to normotension (Table X.02).

Table X.03. CPAP versus no CPAP: Hypertension

Study PMID	Design	Outcome	Analysis (Powered?[75])	Followup Duration (mo)	Arm	CPAP Compliance	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI)[76]
Barbé 2012 22618923[89]	RCT	Incident HTN	ITT	48	CPAP (ITT)	59%[77]	19.8 (15.3, 24.2) events/100 p-yr [75/167(44.9)]	Rate ratio 0.89 (0.65, 1.21) NNT 48 (8, NNH 12)
			Compliant analysis		CPAP (compliant[78])	100% (as treated)	17.6 (12.4, 22.8) events/100 p-yr [44/98 (44.9)]	Rate ratio 0.79 (0.55, 1.14)
			(No)		No CPAP		22.2 (17.5, 26.9) events/100 p-yr [86/183(47.0)]	
Huang 2015 25125635[92]	RCT	Resolution of HTN	As-treated	36 (median)	CPAP	100% (as treated)	25/36(69.4)	**OR 2.98 (1.14, 7.81)[79]**
			(No)		No CPAP		16/37(43.2)	NNT 4 (2, 23)

Statistically significant results are in bold font.

Abbreviations: 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.

[75] Explicitly powered for this outcome?

[76] 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).

[77] Among analyzed participants without hypertension at study baseline.

[78] ≥4 hours/night.

[79] Calculated.

Diabetes Mellitus

Three studies (two RCTs[13, 92] and one NRCS[103] 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)

Both RCTs (Huang 2015 and SAVE) are described above, in the section *Mortality and Cardiovascular Outcomes*. They were both deemed to be at moderate risk of bias related to lack of participant and clinician bias; although, outcome assessors were blinded. Neither trial was explicitly powered for DM-related outcomes. Both analyses of incident DM were restricted to trial participants without DM at enrollment.

The 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 SAVE trial found similar estimates in its ITT and CPAP compliers analyses (see Table X.04). The study did not report an analysis of the association between compliance and risk of DM.

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.[103] Thus, the study was analyzed based on "as-treated" groups. The NRCS may have been subject to uncontrolled biases related to self-selection of CPAP use and compliance. 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 noncompliant 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 a nonsignificant 0.85 (95% CI 0.61 to 1.18). Adding in the adjusted NRCS resulted in a slightly stronger summary estimate of 0.75 (95% CI 0.51 to 1.08), nominally favoring CPAP (Figure X.01). The studies were statistically homogeneous. In the sensitivity analysis of progressively borrowing information from the NRCS to the summary estimate of RCTs, there is a progressive reduction in the effect size, without achieving statistical significance (Appendix Figure D-5).

Heterogeneity of Treatment Effect (DM)

No study reported subgroup effects (or interactions with other factors). No difference was apparent between the as-treated 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).

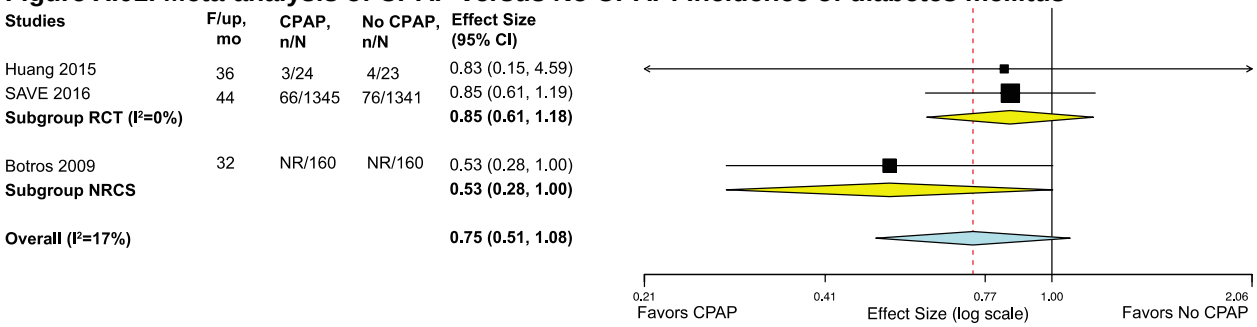
Applicability (DM)

The RCT evidence was mostly applicable to people who would be eligible for enrollment in the SAVE trial, i.e., adults with severe OSA (ODI ≥12) and preexisting CAD or CeVD. Although, it is not evident that in terms of risk for DM, participants in this study substantively differed from the general population of adults with OSA. 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

Two RCTs provide low SoE that CPAP does not affect the risk of incident type 2 DM (Table X.02). The trials were at moderate risk of bias and provided an imprecise estimate of effect (summary effect size 0.85, 95% CI 0.61 to 1.18). Under an assumption of a control rate of 9.9 percent, the summary NNT was 74 (95% CI 28 to −63). 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. It would not be unexpected for future evidence to alter these conclusions.

Figure X.01. Meta-analysis of CPAP versus No CPAP: Incidence of diabetes mellitus



Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, I^2 = measure of statistical heterogeneity ranging from 0% (none) to 100%, NA = not applicable, NR = not reported, NRCS = nonrandomized comparative study, RCT = randomized controlled trial, SAVE = Sleep Apnea cardioVascular Endpoints trial.

Table X.04. CPAP versus no CPAP: Diabetes mellitus

Study PMID	Study Design	Analysis (Powered?[80])	Timepoint (mo)	Arm	CPAP Compliance	n/N (%)	Effect size (95% CI) Calculated NNT (95% CI)[81]
Huang 2015 25125635[92]	RCT	As-treated	36 (median)	CPAP	100% (as treated)	3/24 (12.5)	OR 0.83 (0.15, 4.78)[82]
		(No)		No CPAP		4/23 (17.4)	NNT 20 (4, NNH 6)
SAVE 27571048[13]	RCT	ITT	44.4 (mean)	CPAP	42% overall	66/1345 (4.9)	HR 0.85 (0.61, 1.19) NNT 132 (41, NNH 107(
		(No)		No CPAP		76/1341 (5.7)	[Compliant analysis: HR 0.77 (0.48, 1.27)]
Botros 2009 19958890[103]	NRCS	As-treated	32.4 (mean)	CPAP	100% (as treated)	NR/160	**Adj HR 0.53 (0.28, 0.99)**
		(No)		No CPAP		NR/106	

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.

[80] Explicitly powered for this outcome?

[81] 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).

[82] 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, 97, 118, 120] 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 primarily a trial comparing CPAP with an oral appliance, which included a placebo (sham) oral appliance group.[97] Here we focus on only the comparison of CPAP versus no CPAP. The study reported an intention-to-treat (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 clinical 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 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 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 X.05). 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[121]), none of the effects on depression in the SAVE trial were clinically significant. Studies 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 compliance 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 (Figure X.02). 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

Three of the trials reported on anxiety scores. Aarab 2017 used the SCL-90-R anxiety component. The PREDICT and SAVE trials evaluated the HADS test. Similar to assessment of depression symptoms, the two 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 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 three RCTs all reported changes in anxiety scores (Table X.05). 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[121]), none of the effects on anxiety in the SAVE trial were clinically significant. As for anxiety severity, all trials were conducted as ITT analyses. Aarab 2017 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 −0.11 (95% CI −0.18 to −0.03), which may be interpreted as a small effect size (Figure X.03). 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 analysis of patients with depression symptoms 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).

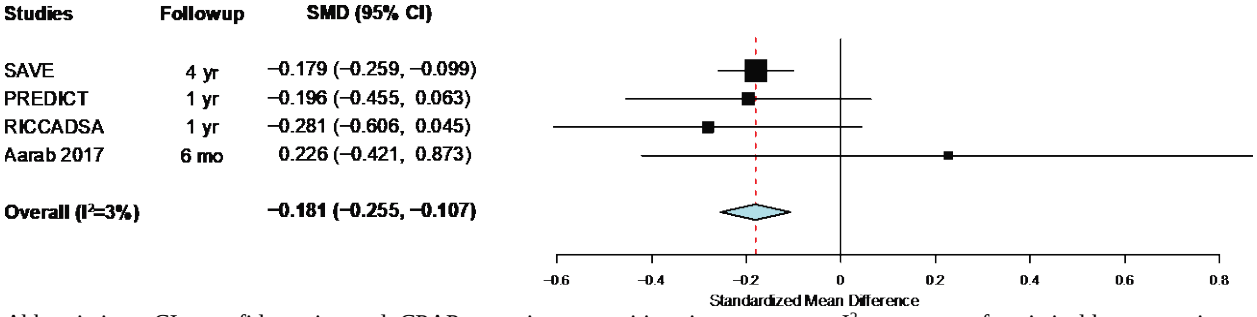
Summary of Effect of CPAP on Mental Health Symptoms

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 X.02). 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 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. It would not be unexpected for future evidence to alter these conclusions.

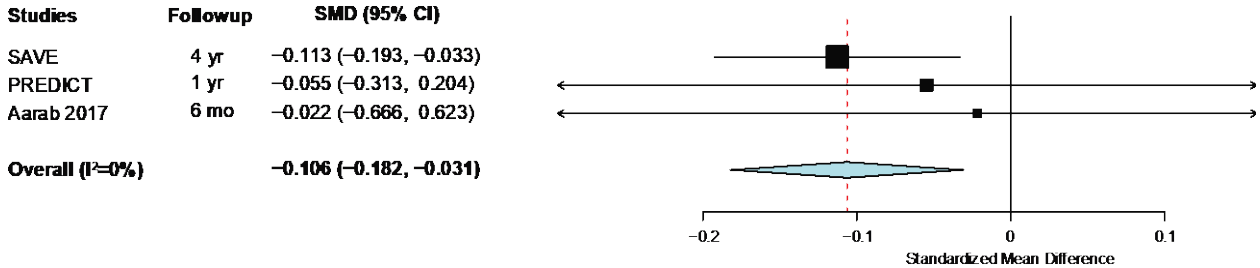
A single trial provides insufficient evidence regarding the effect of CPAP on depression scores specifically among patients with baseline depression symptoms.

Figure X.02. Meta-analysis of CPAP versus No CPAP: Depression scores, standardized mean differences



Abbreviations: CI = confidence interval, CPAP = continuous positive airway pressure, I^2 = 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.

Figure X.03. Meta-analysis of CPAP versus No CPAP: Anxiety scores, standardized mean differences



Abbreviations: CI = confidence interval, CPAP = continuous positive airway pressure, I^2 = 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.

Table X.05: CPAP versus no CPAP: Depression and anxiety scores[83]

Study PMID	Outcome	Scale (Direction) [MCID]	Followup Duration	Arm	CPAP Compliance	N Analyzed (Baseline / Followup)	Mean Diff (SD), Within-Arm	Net Diff (95% CI), Between-Arm
SAVE 31312807[118]	HADS: depression component	0-21 (lower better) [1.9[12184]]	4 yr	CPAP	42% overall	1341/1220	−0.8 (4.0)	**−0.8 (−1.0, −0.5)[85]**
				No CPAP		1336/1190	−0.1 (3.8)	
PREDICT 25172769[93]	HADS: depression component	0-21 (lower better) [1.9[12186]]	1 yr	CPAP	NR	123/114	−0.7 (3.0)	−0.4 (−1.0, 0.3)
				No CPAP		124/116	−0.2 (3.1)	
RICCADSA 30130421[120]	Zung SDS: depression (all participants)	25-125 (lower better) [NR]	1 yr	CPAP	60%	69/69	−3.2 (8.9)	−2.6 (−5.6, 0.4)[87]
				No CPAP		78/78	−0.6 (9.5)	
	Zung SDS: depression (depression at baseline)	25-125 (lower better) [NR]	1 yr	CPAP	60%	30/30	−9.2 (6.8)	**−7.9 (−11.5, −4.3)[88]**
				No CPAP		26/26	−1.3 (7.0)	
Aarab 2017 28083705[97]	SCL−90−R: Depression	NR (lower better) [NR]	6 mo	CPAP	83% of nights[89]	18/18	−3.1 (15.9)	3.6 (−6.5, 13.7)[90]
				No CPAP		19/19	−6.7 (15.3)	

[83] All were randomized controlled trials, analyzed by intention-to-treat, and were not explicitly powered for mental health outcomes.

[84] Per Lemay et al. 2018 in patients with cardiovascular disease.

[85] Adjusted for baseline score

[86] Per Lemay et al. 2018 in patients with cardiovascular disease.

[87] Calculated

[88] Calculated

[89] Self-report. Study included frequent visits (initially about every 2 weeks).

[90] Calculated

Study PMID	Outcome	Scale (Direction) [MCID]	Followup Duration	Arm	CPAP Compliance	N Analyzed (Baseline / Followup)	Mean Diff (SD), Within-Arm	Net Diff (95% CI), Between-Arm
SAVE 27571048[13]	HADS: anxiety component	0-21 (lower better) [1.7[121][91]]	4 yr	CPAP	42% overall	1341/1220	−0.8 (3.6)	**−0.4 (−0.6, −0.2)[92]**
				No CPAP		1336/1190	−0.4 (3.5)	
PREDICT 25172769[93]	HADS: anxiety component	0-21 (lower better) [1.9 [121][93]]	1 yr	CPAP	NR	123/114	−1.2 (3.7)	−0.2 (−0.9, 0.5)
				No CPAP		124/117	−1.0 (3.6)	
Aarab 2017 28083705[97]	SCL−90−R: anxiety	NR (lower better) [NR]	6 mo	CPAP	83% of nights[94]	18/18	−1.7 (9.4)	−0.2 (−0.9, 0.5)
				No CPAP		19/19	−1.5 (8.6)	

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, RICCADSA = Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA trial, SAVE = Sleep Apnea cardioVascular Endpoints trial, SCL−90−R = Symptom Checklist-90-Revised, SD = standard deviation, SDS = Self-rating Depression Scale.

[91] Per Lemay et al. 2018 in patients with cardiovascular disease.

[92] Adjusted for baseline score

[93] Per Lemay et al. 2018 in patients with cardiovascular disease.

[94] Self-report. Study included frequent visits (initially about every 2 weeks).

Cognitive (Executive) Function

Five studies (four RCTs[88, 91, 93, 96] and one NRCS[107]) 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.[93] 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 **APPLES** trial (Apnea Positive Pressure Long-term Efficacy Study) randomly allocated participants to active (n=556) or sham (n=542) CPAP for 6 months.[91] The study reported an intention-to-treat (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 compliant (≥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.[88] The study reported an intention-to-treat (ITT) analysis. Participants were included if they had "moderately severe" OSA (AHI between 10 and 30), without severe daytime sleepiness. 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 clinical 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).

Wu 2016 randomized participants to CPAP (n=68) or no CPAP (n=68) for 6 months.[96] Participants were included if they had an AHI ≥15. The study reported an intention-to-treat (ITT) analysis. The participants had particularly "severe" OSA, with a mean baseline AHI was 61.0; 90 percent were male. The mean age of participants was 49.6 years (range 30 to 65 years) and patients were mostly obese, with a mean BMI of 28.0. The study did not report power calculations. The study did not report compliance 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.[107] The study implicitly conducted an ITT analysis, as it was 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 moderate risk of bias, primarily based on their use of a regression model for adjustment of baseline confounders.

In brief, four studies compared treatment with CPAP with no treatment and the APPLES trial compared real and sham CPAP. All studies were evaluated as ITT 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

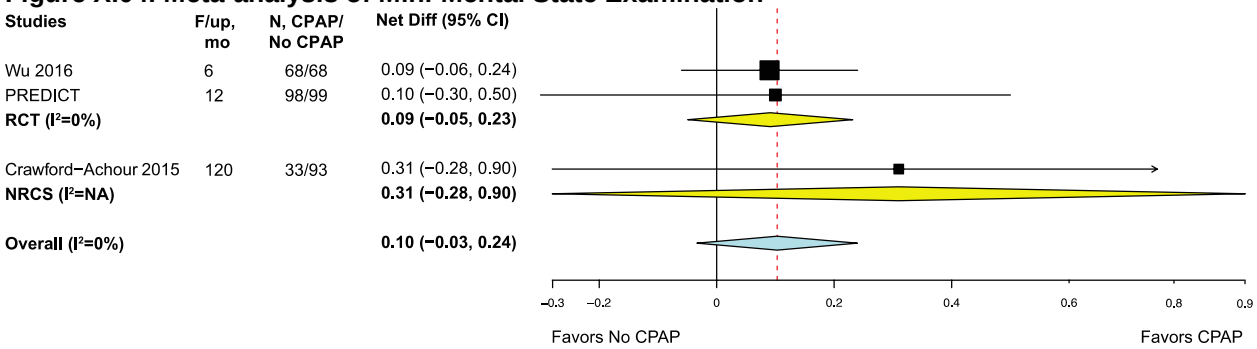
Both the Wu 2016[96] and PREDICT 2014[93] RCTs and the NRCS (Crawford-Achour 2015[107]) evaluated participants with MMSE at 6 months, 12 months, and 10 years, respectively. The MMSE is used to test cognitive function and screen for cognitive loss across the domains of orientation, attention, calculation, 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.[122] 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.[123] However, it is not clear whether changes in MMSE scores is a validated measure of cognitive decline in cognitively normal adults.[124] 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 compliance rates. The NRCS reported 100 percent compliance (≥4 hr/night), but did not provide an explanation for the unusually high compliance rate except that compliance was self-reported.

Meta-analysis of the two RCTs (Figure X.04) yielded a nonsignificant difference in changes in MMSE with or without CPAP at 6 or 12 months, nominally favoring CPAP (net difference 0.09 points, 95% CI −0.05 to 0.23). However, the difference was considerably less than the minimal clinically important difference (MCID) of 1 to 3 points. Inclusion of the 10 year followup NRCS minimally altered the summary estimate to a 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 NRCS evidence (to RCT-only evidence) has very little effect on the summary net difference or its precision.

Figure X.04. Meta-analysis of Mini-Mental State Examination



Abbreviations: CI = confidence interval, CPAP = continuous positive airway pressure, F/up = followup, I^2 = measure of statistical heterogeneity ranging from 0% (none) to 100%, NA = not applicable, Net Diff = net difference.

Trail Making Test B

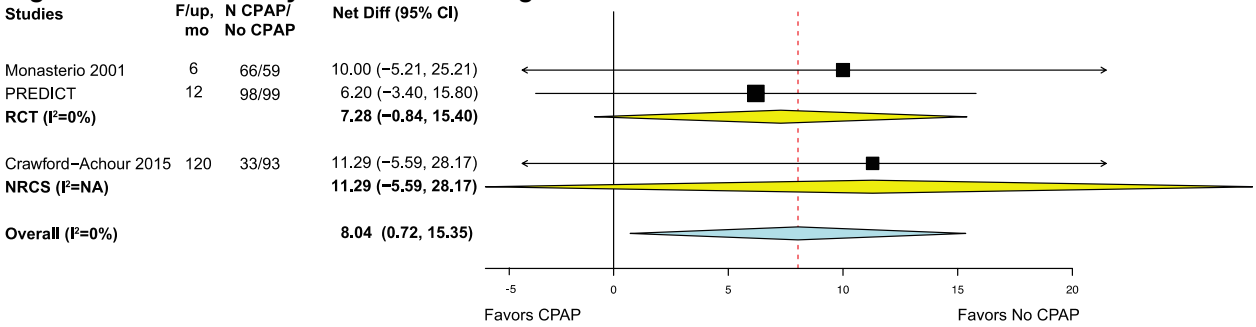
The Monasterio 2001[88] and PREDICT 2014[93] RCTs and the NRCS (Crawford-Achour 2015) evaluated participants on the TMT-B, which involves connecting 25 numbers in circles in ascending order as quickly as possible. In Part B, the circles alternate between numbers and letters (i.e., 1-A-2-B-3-C, …). Results are reported as the number of seconds required to complete the task, so lower scores indicate better cognitive function.[125] Based on a study of patients with CVD (advanced heart failure), a 32-second increase in time to complete the task was provided as a meaningful cognitive decline.[126]

In the PREDICT trial and the Monasterio 2001 RCT, participants in both the CPAP and no CPAP groups had improvements in the TMT-B scores after 12 or 6 months, respectively. The difference between groups were nonsignificant and nominally favored the no CPAP group (PREDICT: net difference 6.2 seconds, 95% CI −3.4 to 15.8; Monasterio 2001: net difference 10, 95% CI −5 to 25). 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), nominally favoring no CPAP.

As noted above, PREDICT did not report on compliance rates and the Crawford-Achour 2015 NRCS had unusually high compliance of 100 percent. The Monasterio 2001 RCT reported 64 percent compliance. None of the studies reported as-treated/compliant user analyses.

Meta-analysis of the two RCTs (Figure X.05) yielded a nonsignificant difference in changes in TMT-B with or without CPAP at 6 or 12 months, nominally 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.

Figure X.05. Meta-analysis of Trail Making Test B



Abbreviations: CI = confidence interval, CPAP = continuous positive airway pressure, F/up = followup, I^2 = measure of statistical heterogeneity ranging from 0% (none) to 100%, NA = not applicable, Net Diff = net difference.

Other Cognitive Tests

The results of all test and subtests are in Appendix D Table 1. Across five 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).[91, 127] 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 three studies each (two RCTs and one 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 X.02). 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. It would not be unexpected for future evidence to alter these conclusions.

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

Six RCTs[13, 90, 93, 95, 98, 128] and one NRCS[106] reported on quality of life (QoL) based on a variety of measures. A seventh RCT reported on functional status outcomes, based on the Functional Outcomes of Sleep Questionnaire (FOSQ).[88] 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 (for example, that the measure covers the "multidomain concept that represents the patient's general perception of the effect of illness and treatment on physical, psychological, and social aspects of life," as defined by the Food and Drug Administration).[129]

Four of the RCTs (MOSAIC, PREDICT, RICCADSA, SAVE) reported CVD outcomes and are described in the section, above, *Mortality and Cardiovascular Outcomes*. In brief, 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.

Monasterio 2001 also reported cognitive function outcomes and is described in the section, above, *Cognitive (Executive) Function*.[88] Briefly, the RCT included participants with AHI between 10 and 30 without severe daytime sleepiness; the study 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.[95] 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 compliant (≥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.[98] The study 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 compliant (≥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 compliance over 2 years.[106] 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). Thus, this study conducted an "as-treated" analysis based on actual use of CPAP. 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 moderate 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), but 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 analyses of CPAP prescription. None of the studies was explicitly powered for QoL or functional status outcomes.

Effect of CPAP on Quality of Life and Functional Status

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

Short Form

The 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.[130] 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.[57, 131]}

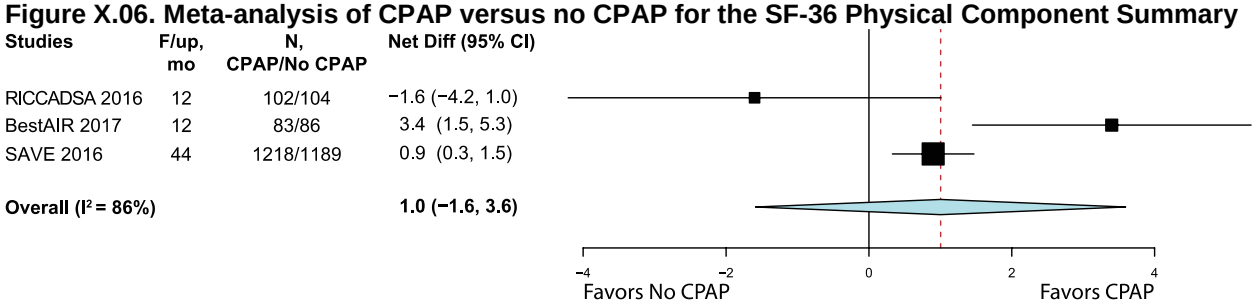
Physical Component Summary

Three RCTs evaluated QoL based on the SF-36 Physical Component Summary at either 1 year (RICCADSA and BestAIR) 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%). The three trials were inconsistent in that SAVE and BestAIR reported a significantly higher net improvement with CPAP,[13, 98] while RICCADSA reported a nonsignificantly worse improvement with CPAP.[128] Meta-analysis of these three studies was most consistent with the largest study (SAVE), yielding a statistically significant 1.0 point (95% CI −1.6, 3.6) higher net improvement for CPAP when compared to no CPAP (Figure X.06). However, the RCTs had statistically heterogeneous findings and the summary estimate is likely not clinically significant (<4 points).

Compliance with CPAP among the three RCTs was 42 percent (SAVE), 52 percent (BestAIR), and 60 percent (RICCADSA). SAVE did not report an analysis among CPAP compliers. BestAIR found no correlation between CPAP compliance 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 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 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); although this P value does not account for multiple 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), consistent with the adjusted ANCOVA p value.

Figure X.06. Meta-analysis of CPAP versus no CPAP for the SF-36 Physical Component Summary



Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, I^2 = measure of statistical heterogeneity ranging from 0% (none) to 100%, 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.

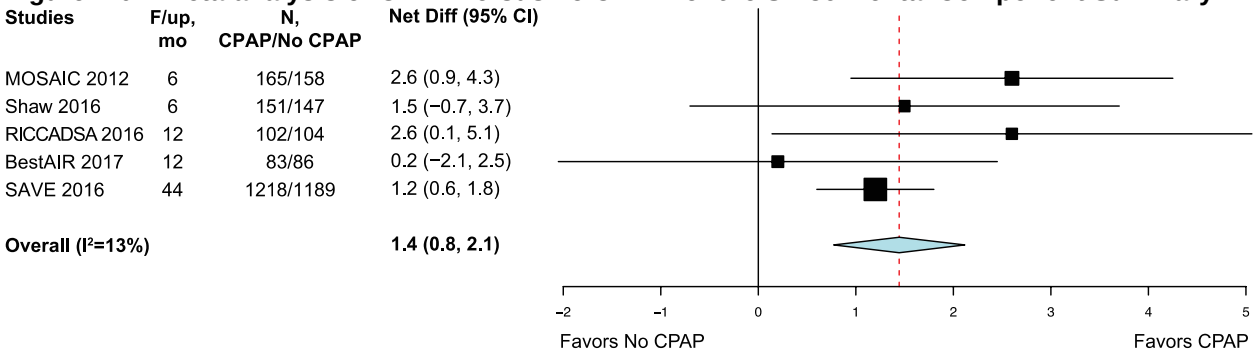
Mental Component Summary

Five RCTs evaluated QoL based on the SF-36 Mental Component Summary at 6 months (MOSAIC and Shaw 2016),[90, 95] 1 year (RICCADSA and BestAIR),[98, 128] or a mean of 3.7 years (SAVE).[13] 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 X.07). The difference is likely not clinically significant (<4 points).

Compliance with CPAP among the five RCTs was about 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 compliers. The MOSAIC trial reported a significantly larger effect among compliant (> hr/night) CPAP users than noncompliant users (4.75 vs. 1.64, P for interaction 0.02). BestAIR found no correlation between CPAP compliance 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 compliant subgroup, but did not report further details including whether the difference was significantly different than among noncompliant 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 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). 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 X.07. Meta-analysis of CPAP versus no CPAP for the SF-36 Mental Component Summary



Abbreviations: 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 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.[132]

MOSAIC and SAVE both reported a small, not clinically significant, net difference of 0.02 on the EuroQol-5D scale, favoring CPAP use over no CPAP use. 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). Neither study compared ITT and CPAP compliers analyses.

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.[133] The index ranges from 1 to 7 with *higher* scores indicating better QoL. The MCID is about 1 to 2 points.[58]

MOSAIC and PREDICT reported statistically significantly higher net improvements on the SAQLI scale, of 0.6 (95% CI 0.4 to 0.8) points[90] and 0.4 (95% CI 0.2, 0.6),[93] which were likely not clinically significant. Both trials analyzed most randomized participants (MOSAIC 84%, PREDICT 86%). Neither study compared ITT and CPAP compliers analyses.

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;[64] although we found no studies that have validated it against established OSA-related hrQoL outcomes. The scale ranges from 0 to 120 points, with *higher* scores indicating better functional status. The MCID is about 17 to 20 points.[134]

Monasterio 2001 evaluated functional status using a Spanish version of the original FOSQ-30.[88] 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"[135] 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, nominally favoring CPAP: net difference 3.0 points (95% CI −3.6 to 9.6). The nominal difference is not likely to be clinically significant. The study reported no significant difference in outcome between the CPAP compliers (64%) and noncompliant effects.

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 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.[106] 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 compliance 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 compliance. Two RCTs suggested possible larger effects among CPAP compliant users than noncompliant 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

Three RCTs provide low SoE that CPAP does not have a clinically significant effect on the Physical Component Score of the SF-36 (Table X.02). Although imprecise, all estimates of effect were less than the clinically significant threshold of about 4 to 7 points (summary net difference

1.0 (95% CI −1.6 to 3.6). Studies found no association between CPAP compliance 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 compliance with CPAP was associated with the Mental Component Score. The measure has not been validated in an OSA population.

Similarly, two RCTs provide low SoE that CPAP does not have clinically significant effects on either the EuroQoL-5D and the SAQLI measures of QoL. The measure has not been validated in an OSA population.

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. It would not be unexpected for future evidence to alter these conclusions.

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

Sexual Function

Three studies (one RCT[97, 136] and two NRCS[104, 112] 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.[136]

Budweiser 2013 was an observational study derived from a cohort of men who had had diagnostic polysomnography for suspected OSA.[104] Although reporting was unclear, the study apparently was an "as treated" analysis of the comparison of actual users versus nonusers (including noncompliant) with CPAP. Among these 401 men, 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 related to inadequate description of assignment to CPAP, that the study likely combined patients not prescribed CPAP with those who were noncompliant with CPAP, and inadequate reporting of the adjusted analysis results.

Jara 2018 was an observational study that compared CPAP use to no use based on compliance; thus, as an "as-treated" analysis.[112] 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 compliance (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 compliant users of CPAP with noncompliant/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[136]) used the sexual/social dissatisfaction measure on the Sleep Disorders Questionnaire (SDQ; range 0 to 5 with *lower* scores indicating better sexual function).[137] 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 compliance 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).[138] 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, 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[112] created a sexual function measure based on two sex-specific questions from the Symptoms of Nocturnal Obstruction and Related Events (SNORE-25) instrument.[40] The validity of this de novo specific submeasures 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 (compliant) CPAP users had improved sexual function after 12 months (change in score −0.7) units compared with a marginal improvement in the (noncompliant) non-CPAP 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.[112]

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.[104]

The Jara 2018 NRCS reported subgroup multivariable-adjusted results separately for men and women.[112] 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 clinical outcomes comparing CPAP with other active treatments are relatively sparse. None was designed as a noninferiority or equivalence trial regarding clinical 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 clinical outcomes have not been reported.
- There is low SoE that functional status remains similar between patients prescribed fixed CPAP or autoCPAP. Other long-term clinical outcomes have not been reported.

Four studies that reported long-term clinical outcomes compared CPAP to other active treatments (n=2) or different CPAP devices (n=2) 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 clinical outcomes.

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.

Mental Health (CPAP Versus MAD)

Two RCTs compared CPAP with mandibular advancement devices (MAD),[97, 139] evaluating different measures of depression and anxiety (Table CPAP.MAD.01).

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).[97] 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 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).

De Vries 2019 compared CPAP (n=42) to MAD (n=43) in 85 participants with OSA in an ITT analysis.[139] 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 clinical 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 of the 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 CPAP.MAD.03). 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. It would not be unexpected for future evidence to alter these conclusions.

Quality of Life and Functional Status Outcomes

Two RCTs compared different CPAP modalities and reported on FOSQ, which has been validated against other QoL measures in CPAP-treated patients with OSA.[64] One RCT compared CPAP with MAD and reported FOSQ and various QoL measures, which do not appear to have been validated specifically in adults with OSA.

Comparison of CPAP Modalities

Two RCTs (Bloch 2018 and Kushida 2011) compared different CPAP modalities.[140, 141] Both reported on FOSQ results (Table CPAP.MAD.02). FOSQ has been validated against other QoL measures in CPAP-treated patients with OSA.[64]

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.[140] The study was designed to test equivalence of autotitrated and fixed CPAP for ESS and sleep resistance, but not for clinical 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 autoCPAP had a pressure range from 5 to 15 mbar; fixed CPAP had pressure set at the 90th percentile applied by the autoCPAP device during adaptation. Participants were provided either a Philips Respironics REMstar or a ResMed AutoSet device. 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.[141] 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. Both groups used the REMstar Auto M-Series CPAP device (Philips Respironics). For participants assigned to autotitrated CPAP, the device was set to reduce air pressure during exhalation. For participants assigned to fixed CPAP, the device did not alter pressure during the breathing cycle. 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 17 to 20 points.[134]). Based on these two studies, there is moderate SoE that functional status does not differ in patients prescribed autoCPAP or fixed CPAP (Table CPAP.MAD.03). The moderate SoE suggests that we have moderate confidence that the summary estimates (and their confidence intervals) are close to the true effect. The body of evidence has some deficiencies. Additional evidence would most likely support the current findings, but some doubt remains.

CPAP Versus MAD

One RCT comparing CPAP and MAD (de Vries 2019) reported on QoL and functional status based on the SF-36, EuroQol-5D, and FOSQ in CPAP compared to MAD at 1 year (Table CPAP.MAD.01).[139] For SF-36, we calculated the estimated Physical and Mental Component Summary scores from the reported mean specific component scores.[142, 143]

The SF-36 Physical Component Summary improved in both groups, with a small, not clinically significant, difference between groups, which had imprecise estimates: net difference 2.4 (95% CI −20.7 to 25.5; MCID 4 to 7 points[57]).

The SF-36 Mental Component Summary also improved in both groups, but there was a negligible difference between groups: 0.4 (95% CI −20.2 to 20.9).

Similar results were also found 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 4 to 7 points[57]).

The analysis of overall functional status, as measured with the FOSQ, also found improved functional status scores at 1 year in both groups but a small, not clinically or statistically significant difference between groups: difference between changes in median values −0.1 (MCID about 17 to 20 points.[134]).

The trial did not evaluate subgroup differences.

In summary, based on one small RCT that provided imprecise estimates of differences between interventions, there is insufficient evidence to evaluate the relative effect of CPAP or MAD on QoL or functional status (Table CPAP.MAD.03).

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.[97, 136] The study evaluated SDQ in 18 participants using CPAP and 20 using MAD for 6 months (Table CPAP.MAD.01).[136] 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 CPAP.MAD.03).

Table CPAP.MAD.01. CPAP versus MAD: 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[97]	SCL−90−R: Depression	NR (lower better) [NR]	6 mo	CPAP	18/18	−3.1 (15.9)	−0.8 (−9.4, 7.8)
				MAD	20/20	−2.3 (10.3)	
de Vries 2019, 31596213[139]	HADS: Depression component	0-21 (lower better) [1.9[121,95]]	1 yr	CPAP	39/37	[−1.0]	[1.0], NS
				MAD	40/29	[−2.0]	
	Anxiety						
Aarab 2017 28083705[97]	SCL−90−R: Anxiety	NR (lower better) [NR]	6 mo	CPAP	18/18	−1.7 (9.4)	0.3 (−4.7, 5.3)

[95] Per Lemay et al. 2018 in patients with cardiovascular disease.

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
				MAD	20/20	−2.0 (4.2)	
de Vries 2019, 31596213[139]	HADS: Anxiety component	0-21 (lower better) [1.7[12196]]	1 yr	CPAP	39/37	[−1.0]	[0], NS
				MAD	40/29	[−1.0]	
	QoL/Functional status						
de Vries 2019, 31596213[139]	SF-36, PCS (estimated)[97]	0-100 (higher better) [~4-7]	1 yr	CPAP	39/37	7.2 (51.0)	2.4 (−20.7, 25.5)
				MAD	40/29	4.9 (51.8)	
	SF-36, MCS (estimated)[98]	0-100 (higher better) [~4-7]	1 yr	CPAP	39/37	1.9 (45.5)	0.4 (−20.2, 20.9)
				MAD	40/29	1.6 (46.1)	
	EuroQol-5D	0-100 (higher better) [~18]	1 yr	CPAP	39/37	4.0 (12.9)	−0.8 (−7.6, 6.0)
				MAD	40/29	4.8 (14.7)	
	FOSQ	5-20 (higher better) [~2]	1 yr	CPAP	39/37	[2.4]	[−0.1], NS
				MAD	40/29	[2.5]	
	Sexual function						
Aarab 2017, 28294380[136]	SDQ, sexual/social dissatisfaction	0-5 (lower better), NR	6 mo	CPAP	18/18	0 (0.9)	0.2 (−0.4, 0.8)
				MAD	20/20	−0.2 (1.1)	

Abbreviations: 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, SCL−90−R = Symptom Checklist-90-Revised, SD = standard deviation, SDS = Self-rating Depression Scale, SF = Short Form.

Table CPAP.MAD.02. Comparison of CPAP devices: Functional status (per FOSQ)

Study PMID	Followup Duration	Arm	N Analyzed	Mean Diff (SD), Within-Arm	Net Diff (95% CI), Between-Arm
Bloch 2018 28982804[140]	24 mo	AutoCPAP	113	3.0 (2.4)	−0.4 (−0.9, 0.1)[99]
		Fixed CPAP	95	2.6 (2.2)	
Kushida 2011 21804670[141]	6 mo	AutoCPAP	54/46	1.9 (2.8)	−0.2 (−1.6, 1.1)
		Fixed CPAP	57/47	3.3 (2.2)	

Abbreviations: 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.

[96] Per Lemay et al. 2018 in patients with cardiovascular disease.

[97] Calculated from reported specific component scores.

[98] Calculated from reported specific component scores.

[99] Adjusted difference

Table CPAP.MAD.03. 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	RCT or adj NRCS	0							None
Accidents	RCT or adj NRCS	0							None
Hypertension	RCT or adj NRCS	0							None
Diabetes	RCT or adj NRCS	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	RCT or adj NRCS	0							None
QoL & functional status (CPAP vs. MAD)	RCT	1 (85)	Moderate	N/A	Highly imprecise	Indirect	Single study	Insufficient	No conclusion
Functional status (auto vs. fixed CPAP)	RCT	2 (301)	Low to Moderate	Consistent	Precise	Direct	None	Moderate	No evidence of difference
Sexual function (CPAP vs. MAD)	RCT	1 (38)	Moderate	N/A	Imprecise	Indirect	Single study	Insufficient	No conclusion
Days of work missed	RCT or adj NRCS	0							None

Abbreviations: adj NRCS = adjusted nonrandomized comparative studies, CPAP = continuous positive airway pressure (device), CV = cardiovascular, MAD = mandibular advancement device, N/A = not applicable, QoL = quality of life, RCT = randomized controlled trials, SoE = strength of evidence.

Evaluations of RCT evidence base are in bold font.

Key Question 1: CPAP Adverse Events

Key Points

- Studies that report comparative, long-term clinical outcomes provide insufficient evidence regarding adverse events.
- Based on 854 case reports of adverse events available in the Food and Drug Administration 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.

Evidence

Only two RCTs reported adverse events related to CPAP use.[93, 95] 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 AE.01). 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).[144-146] 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).[93] Shaw 2013 prescribed a similar autotitrating CPAP device (S8 Autoset Spirit II, ResMed), with pressure settings between 5 and 20 cm H_2O.[95]

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 AE.01. 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[95]	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[95]	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[100]	PREDICT[93]	AutoCPAP (S9 Autoset, ResMed)	12 mo	47/140 (33.6)	NR		
	Epistaxis	Shaw 2016[95]	AutoCPAP(S8 AutoSet Spirit II; ResMed)	6 mo	1/151 (0.7)	0/147 (0)	0.7 (−0.6, 2.0)	1.00
Psychological	Anxiety/dyspnea related to CPAP	PREDICT[93]	AutoCPAP (S9 Autoset, ResMed)	12 mo	4/140 (2.9)	N/A		
Musculoskeletal	Pain, musculoskeletal	Shaw 2016[95]	AutoCPAP(S8 AutoSet Spirit II; ResMed)	6 mo	1/151 (0.7)	1/147 (0.7)	0 (−1.9, 1.8)	1.00
Gastrointestinal	Abdominal bloating	PREDICT[93]	AutoCPAP (S9 Autoset, ResMed)	12 mo	4/140 (2.9)	NR		
	Gastrointestinal distress	Shaw 2016[95]	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[101]	PREDICT[93]	AutoCPAP (S9 Autoset, ResMed)	12 mo	33/140 (23.6)	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.

[100] E.g., dry mouth, runny or stuffy nose, sinus problems, nose bleeds

[101] 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 (or similar measures) are valid intermediate or surrogate measures for long-term clinical outcomes.
- 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 clinical 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 a given breathing measure and clinical 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 clinical 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 and clinical outcomes.

We found no study that formally validated changes in breathing measures as intermediate or surrogate outcomes for clinical outcomes. No study reported within-study participant-level correlation analyses between intermediate or surrogate outcomes and clinical 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 polysomnography measures and multiple instances of incomplete reporting of definitions and criteria.

Among the 11 studies that evaluated changes in AHI, polysomnography was 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 (one 1997, two 1999, two 2007, two 2012, one 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, polysomnography was 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 one study, and not reported in the other two studies.

Correlation of Changes in Intermediate or Surrogate Measures With Clinical Outcomes

From the 15 eligible studies, we extracted 50 sets of breathing measure and clinical 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 clinical outcomes and/or multiple breathing measures. However, most comparisons between a given breathing measure and clinical 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 Clinical Event Outcomes

Only three RCTs reported changes in breathing or sleepiness measures in both randomized groups and reported clinical event outcomes (Table Corr.01). As noted, none reported a patient-level correlation between measures and clinical 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,[13] at enrolment, patients, AHI was estimated from a two-channel (oximetry and nasal pressure) home device; apneas and hypopneas 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 clinical 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 outcomes. The SAVE trial also reported change in ESS, which similarly favored CPAP (net change −2.5, 95% CI −2.8 to −2.2), which also did not correspond with the lack of significant effect on clinical outcomes. 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).[147] 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 polysomnography, 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 polysomnography AHI was 20.9, while laboratory AHI was 28.5; statistical significance is not reported. Potentially, thus, participants in the home polysomnography 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 polysomnography 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 possibly correlation between change in AHI (not to mention a single definition of AHI) and clinical outcomes. As shown in Table Corr.01, there was also no evidence suggesting a correlation between ESS and clinical 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 Corr.01).[93] While the risk of composite CV events nominally favored CPAP prescription, 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 Corr.01. Trials reporting both changes in intermediate measures and clinical effect sizes

Study, PMID	Followup Interval	Arm	Measure	Estimated Net Difference (95% CI)	Event	Relative Difference (95% CI)
Corral 2017, 28636405[147]	6 mo	Home vs. lab PSG	ESS	0.7 (−0.2, 1.6) Favor (NS) lab PSG	Accident	1.1% (−3.7, 5.7) Favor (NS) lab PSG
			AHI*	1.4 (−1.2, 4.0) Favor (NS) lab PSG	CV incidence	−0.9% (−6.9, 5.1) Favor home PSG
PREDICT, 25172769[93]	1 yr	CPAP vs. no CPAP	ESS	−2.1 (−3.0, −1.2) Favor CPAP	Composite CV events[102]	HR 0.87 (0.40, 1.88) Favor (NS) CPAP
SAVE, 27571048[13]	3.7 yr (mean)	CPAP vs. no CPAP	ESS	−2.5 (−2.8, −2.2) Favor CPAP	Composite CV events[103]	HR 1.10 (0.91, 1.32) Favor (NS) no CPAP
			AHI	−25.3 (−24.1, −22.2), CPAP group		

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

* Measured differently in two groups. See text.
† 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 Corr.01A).[139, 147-149] Note that, as described in the *Quality of Life and Functional*

[102] Specific event outcomes were rare and effect estimates were imprecise.
[103] Primary outcome. Other event outcomes also statistically nonsignificant.

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 Corr.01B).[139, 147-149] 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

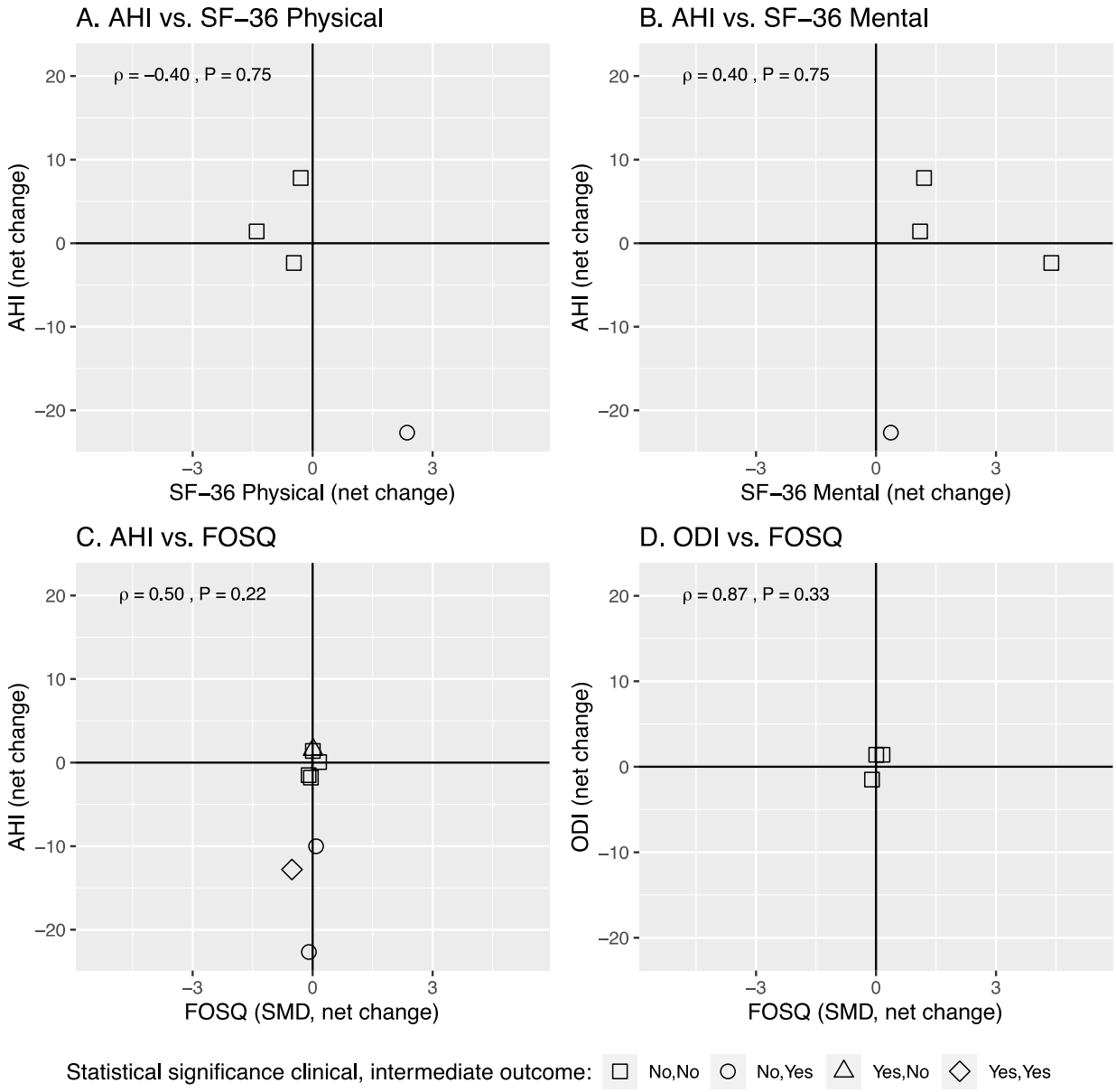
AHI and FOSQ

Eight studies reported data to allow a comparison between changes in AHI and the FOSQ measure of functional status (Figure Corr.01C).[88, 139-141, 147, 150, 151] Note that FOSQ has been validated against other QoL measures in CPAP-treated patients with OSA.[64] 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 Corr.01D).[140, 147, 150] 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 Corr.01. Correlation between changes in breathing measures and clinical outcomes



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

From the 12 eligible studies, we extracted 85 sets of ESS measurements and clinical outcome pairs. Studies commonly reported multiple clinical outcomes. However, most comparisons between a given breathing measure and clinical 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 Corr.02A).[13, 93, 139] 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 Corr.02B).[13, 93, 139] 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 Corr.02C).[13, 98, 139, 147-149] 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

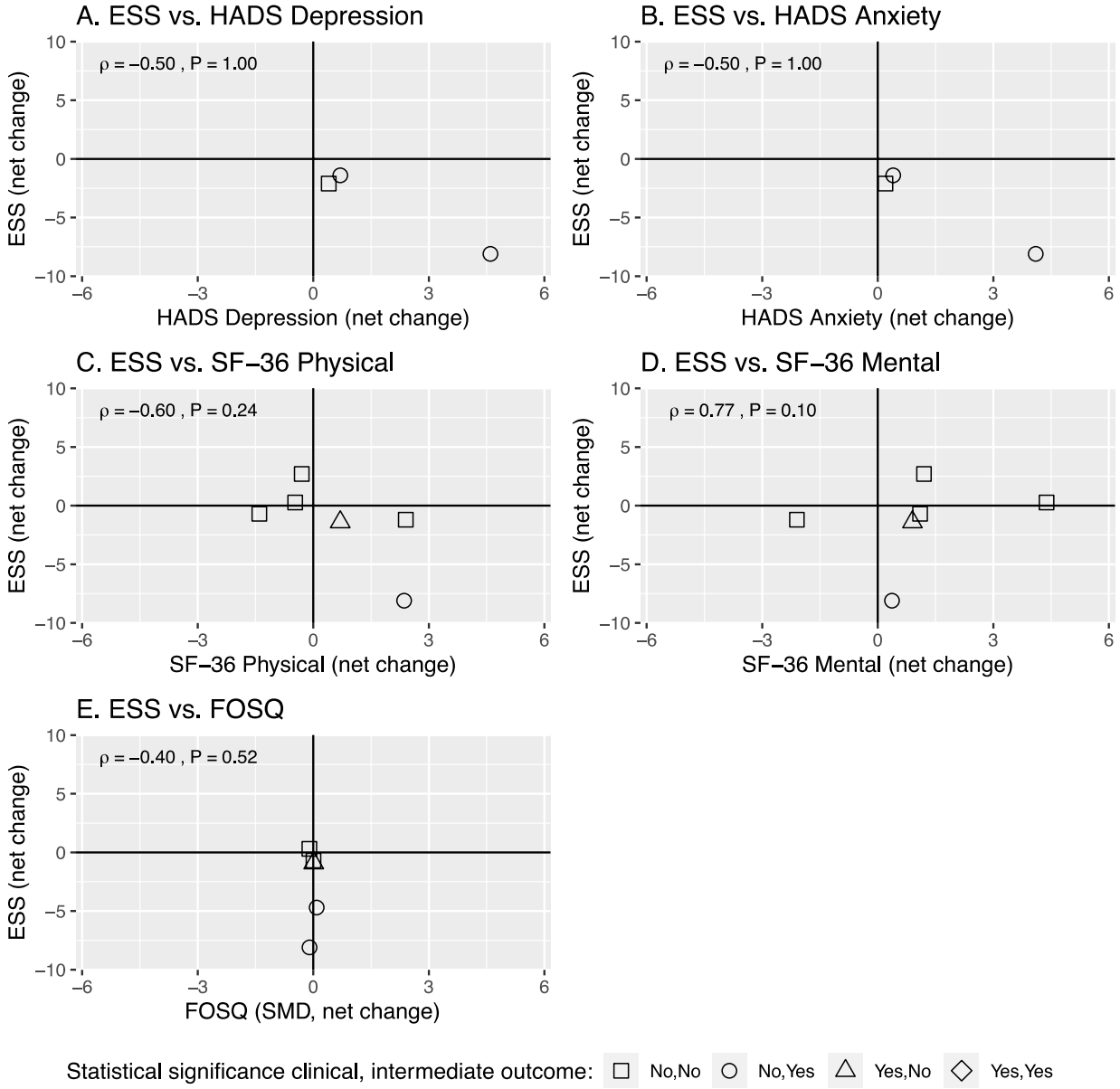
The same six studies reported data on changes in ESS and the SF-36 Mental Component Summary score (Figure Corr.02D).[13, 98, 139, 147-149] Across studies there was a moderate "negative" correlation (correlation coefficient = 0.77), consistent with larger improvements (reductions) in ESS correlating with 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 Corr.02E).[88, 139-141, 147] 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 Corr.02. Correlation between changes in sleepiness scores and clinical outcomes



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

Abbreviations: ρ = correlation value, FOSQ = Functional Outcomes of Sleep Questionnaire, HADS = Hospital Anxiety and Depression Score, Mental = Mental Component Score, P = P-value (statistical significance), Physical = Physical Component Score, SF = Short Form, SMD = standardized mean difference.

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

Overall, the evidence base neither supports nor refutes whether commonly used measures (AHI, ODI, ESS) are valid intermediate or surrogate measures for long-term clinical outcomes.

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 clinical outcome of interest. Thus, no conclusions can be made regarding whether any measures may be actual intermediate or surrogate outcomes for clinical 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)[152] but not concomitant reductions in risk of composite CV events. Consistent with numerous RCTs that evaluated AHI,[12] incidentally, the SAVE trial found that CPAP reduced AHI by a large degree (−25.3), but nonsignificant effects on composite CV events, mortality, and other clinical outcomes. However, none of these findings provide evidence whether AHI or ESS are valid surrogate or intermediate measures for clinical 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 or fully, exhibited.[153, 154] 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 clinical 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.Corr.03. 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 clinical 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).[155, 156] 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.[153, 154, 157, 158] 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 $\frac{NDE\times(NIE-1)}{NDE\times NIE-1}$.

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.

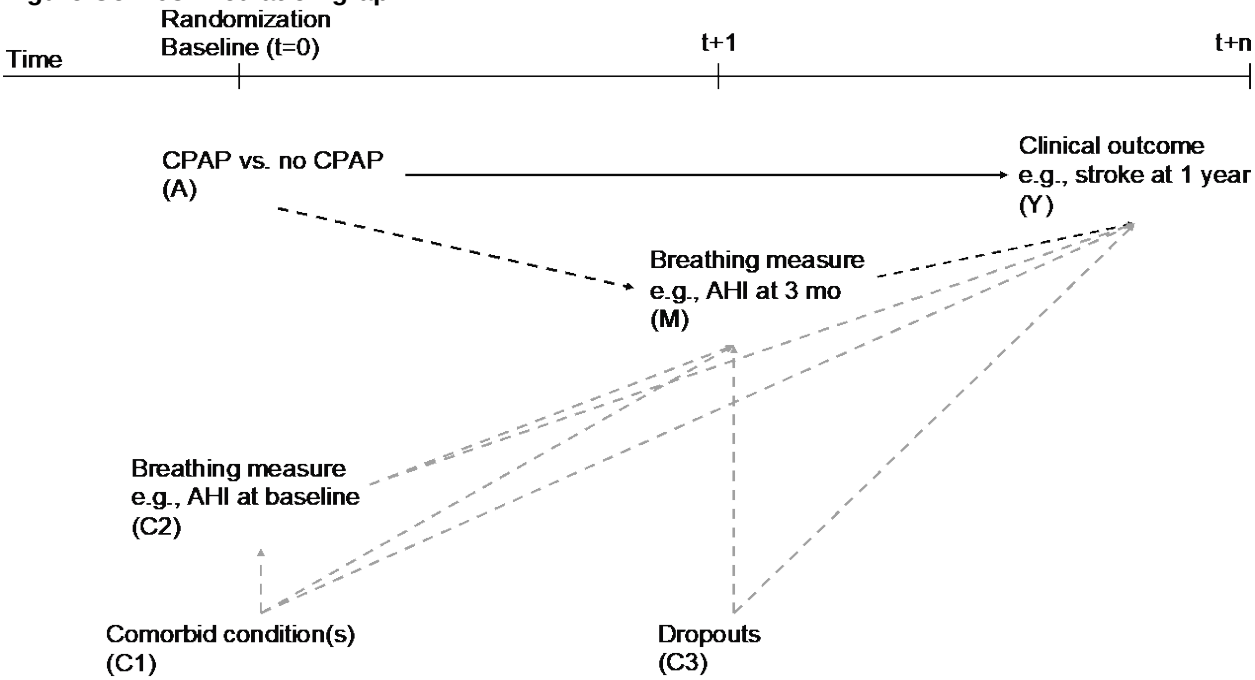
Surrogacy Analyses

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 clinical outcome. It would not be necessary for AHI to be a mediator of the clinical outcome, as long as the change in AHI tracks with, and therefore predicts, the effect of CPAP on the clinical 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.[159] 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 Corr.03. Mediation graph



Directed acyclic graph (DAG) of mediation of a breathing measure on a clinical 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, and many of the studies, though of at least fair methodological quality, provided largely imprecise estimates of comparative effectiveness, resulting in generally low strength of evidence (SoE) conclusions. The implication of low SoE is that "we have limited confidence that the estimate of effect lies close to the true effect for this outcome. The body of evidence has major or numerous deficiencies (or both). We believe that additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect."[30]

Definitions of Breathing Measures Findings

Across all studies included in this systematic review, studies were highly inconsistent in the criteria used to define polysomnography measures (apnea, hypopnea, and oxygen desaturation), 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 polysomnography 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. Furthermore, 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), further complicating interpretation of how AHI (etc.) 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 ODI 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. 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. It would not be unexpected for future evidence to alter these conclusions. Randomized controlled

trials (RCTs) do not demonstrate that CPAP affects all-cause mortality or various cardiovascular (CV) outcomes, clinically important changes in psychosocial measures, or other clinically important outcomes. Inclusion of nonrandomized comparative study (NRCS) evidence, however, suggests that CPAP reduces the risks of all-cause mortality (low SoE), but does not change other conclusions.

Regarding all-cause mortality, the RCTs and NRCSs generally agree in their effect estimates (which ranged from 0.48 to 0.91, excluding two small, imprecise studies with even smaller effect sizes, both favoring CPAP). Similarly, the NNTs suggested by the RCTs ranged from 47 to 426; 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 the largest RCT reporting all-cause mortality (SAVE), 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.

The RCTs were all underpowered to detect differences in all-cause mortality rates (i.e., they were statistically nonsignificant), while the NRCSs all found statistically significant effects. While one cannot rule out selective outcome reporting (publication bias), particularly among the NRCSs, as an explanation for the consistent findings, it can be argued that a contradictory finding (of no effect or of increased death with CPAP) would be of particular interest for publication. The NRCSs mostly had greater power to find a difference in risk of death, related to higher event rates (percent who died) due to either older age (e.g., ≥60 or ≥80 years), longer-term followup (up to 11 years), or particularly large sample size (N >25,000). Thus, it is plausible that the conclusion of a reduction in risk of death may be most applicable to older adults (at increased risk of dying) and considerations of longer-term followup.

Combining RCTs and NRCSs, there is evidence (though of low strength) that CPAP reduces the risk of all-cause mortality, but there is mostly not evidence 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. 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). Where we concluded that CPAP may have no effect, with low SoE, it is more accurate to interpret this as a finding that there is no evidence of an effect of CPAP rather than that there is definitive evidence that CPAP has no effect on outcomes. Thus, it is possible that the current evidence base does not reveal real differences in risk of specific outcomes with CPAP use. These (not statistically significant) differences in risk of specific outcomes may, then, 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, among RCTs that reported data, compliance rates ranged from 38 to 67 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 compliant users analyses than the intention-to-treat (ITT) 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. This may suggest that CPAP is ineffective, since its effect is equivalent among users and nonusers. But it is unclear whether the lack of significant differences between compliant 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 compliant and noncompliant 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 (maybe such as increased communication with the sleep clinic) or that even the "low dose" of CPAP achieved by noncompliant users is effective. However, such explanations are just conjectures and would need to be explored critically.

In contrast to findings about the effect of compliance regarding CV event outcomes, there is conflicting evidence across the RCTs that evaluated the Mental Component Score of the Short Form (SF) 36. Two of four RCTs that compared effect of CPAP in compliant and noncompliant users found stronger beneficial effects among the compliant CPAP users. Although, in the one trial that provided sufficient data (MOSAIC), the difference was not clinically significant (4.75 vs 1.64 points).

Complicating the comparison of compliant and noncompliant participants (and the comparison of studies with ITT analyses and NRCSs of compliant vs. noncompliant patients), is that compliance is not a random event. People who do or do not comply may differ in how "effective" CPAP is for them. However, if there is a difference, likely it has more to do with how well CPAP works for immediate symptoms (e.g., sleepiness, snoring) than for long-term clinical outcomes. It is likely that the NRCSs and as-treated analyses were not able to fully control for the inherent differences between CPAP users and non-users/non-compliant individuals, potentially biasing the studies toward increased effectiveness of CPAP.

For non-CV outcomes, the RCTs provide low SoE that CPAP improves measures of depression and anxiety symptoms, executive cognitive function, and quality of life (QoL), but, importantly, these changes are small and not clinically significant. Furthermore, no studies evaluated clinical 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 clinical outcomes; thus, most conclusions were unchanged when also considering the NRCSs.

For numerous clinical outcomes of interest, there is insufficient evidence, primarily because only a single study evaluated a given outcome or 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 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 (Nakamura 2009) reported adjusted hazard ratios (HR) among different subgroups by AHI,[101] 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, 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,[91, 127] but no differences were noted. Another NRCS (Bjornsdottir 2015) reported similar effects of CPAP on QoL as measured by the Short Form 12 (SF-12) in subgroups based on sleepiness symptoms (ESS ≥10 vs. <10).[106]

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 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 to 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 clinical 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 (MAD) in changes in depression or anxiety symptom scores, and insufficient evidence (based on a single study each) regarding changes in QoL or sexual function measures. Similarly, based on two small studies, no differences were found in OSA-related functional status (by the Functional Outcomes of Sleep Questionnaire [FOSQ]), with moderate SoE. The moderate SoE suggests that we have moderate confidence that the summary estimates (and their confidence intervals) are close to the true effect. The body of evidence has some deficiencies. Additional evidence would most likely support the current findings, but some doubt remains. Further conclusions regarding other clinical outcomes or potential heterogeneity of treatment effect are not feasible, given the limited evidence base.

Intermediate and Surrogate Measures Findings

The large majority of CPAP RCTs did not meet eligibility criteria for this systematic review mostly because they did not report on long-term clinically important 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 these (and related) intermediate measures are valid surrogate or mediator outcomes for the long-term clinical outcomes. To qualify, the intermediate measures had to, at a minimum, correlate with the clinical outcomes. However, we found no study that evaluated whether any intermediate outcome may be a valid surrogate or mediator measure. 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 clinical 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 clinical 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 (when CPAP was being used) over the 3.7 years of followup on CPAP (among those randomized to the device). Despite the large improvement in AHI, CV events and other clinically important outcomes did not improve and, in fact, the hazard ratio for their primary outcome (composite CV events) nominally 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 evidence whether AHI (or ESS) are valid surrogate or intermediate measures for clinical 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 clinical outcome risk reduction). There is 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 clinical outcomes. Numerous other explanations are possible.

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 clinical 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 clinical 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 clinical 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 clinical 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 clinical outcomes.

Strengths and Limitations

CPAP (Strengths and Limitations)

The evidence base examined in this systematic is sparse or of low SoE to establish the long-term effects of CPAP treatment on clinical outcomes in patients with OSA. As noted, for most clinical 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).

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). Separate from whether AHI is a valid measure of OSA severity or whether change in AHI is a valid intermediate or surrogate measure of clinical 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 to 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 relatively greatly affect 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 most (or least) clinically benefit from CPAP (assuming any patients do). Even had there been more (statistical) variability across studies in the effects of CPAP treatment, we would have been unable to meaningfully evaluate whether any of the variability 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. Polysomnography 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.

Related, the high variability 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 (Sleep Apnea cardioVascular Endpoints) trial is the largest RCT to date comparing CPAP to no CPAP with long-term clinical outcomes; it was of modest size, with 2687 participants followed for about 4 years. Even with restriction to patients with underlying CV 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 (hazard ratio = 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 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.

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.[91, 98] While it is apparent that implementing sham CPAP for years in a large group of study participants would be difficult and impractical, 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 clinical 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, with one low risk of bias RCT, six moderate risk of bias RCTs, and four high risk of bias RCTs. 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 compliant 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, Nakamura 2009) 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, effect sizes from NRCSs tended to be fully consistent with those from RCTs; although, NRCSs generally had somewhat stronger effect sizes that were more likely to be statistically significant than RCTs. This was particularly true for studies of all-cause mortality and, to a lesser degree, composite CV outcomes.

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 clinical outcomes is insufficient to make conclusions. Importantly, no comparative studies with long-term clinical outcomes evaluated potential correlations of intermediate/surrogate and clinical 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).[13] 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). Given its relatively large size, conclusions regarding outcomes reported by SAVE (most CV-related outcomes, accidents, incident diabetes, and depression and anxiety) could be considered to be most applicable to the population included in SAVE. However, with the exception of the SF-36 Physical Component Score (with an I^2 of 86%), RCTs were mostly consistent (homogeneous) in their reported outcomes (I^2 <40%), implying no substantive effect size differences across studies (and, thus, included study populations). The NRCSs, likewise, had mostly consistent findings.

This all argues in favor of broad applicability of the conclusions to adults with OSA who are being considered for CPAP treatment. 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

Most conclusions from this systematic review were of low strength of evidence, mostly finding no statistically or clinically significant effects of CPAP on clinical outcomes. The overall lack of evidence supporting an effect of CPAP on long-term clinical outcomes stands in contrast to the substantive effect of CPAP on sleep measures such as AHI and symptoms, as evaluated by ESS.[12] Furthermore, we were unable to make meaningful conclusions regarding the effect of CPAP on numerous clinical outcomes, or regarding the validity of breathing or symptom measures as intermediate or surrogate measures for clinical outcomes. The most important limitations were imprecise (frequently highly imprecise) estimates of effect and sparse evidence. Most of the important clinical outcomes are relatively infrequent events (even among selected populations) that 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 evidence supports a lack 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 QoL were small and below thresholds for clinically meaningful differences in effect.

Thus, overall, the current RCT evidence base provides conclusions based on low strength of evidence) 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. For the most part, the associations found by adjusted NRCSs are consistent with the direction and magnitude of the effects of CPAP in the RCTs. For all-cause mortality, the additional evidence from the NRCSs suggests possible effectiveness of CPAP—albeit with a low strength of evidence, and mostly applicable to older adults (>60) and/or longer-term followup (up to 11 years). Future RCTs are

needed to confirm or refute the conclusions. The trial evidence supports a lack of clinically meaningful effect on psychosocial measures.

The generally low SoE regarding the use of CPAP to prevent long-term clinical outcomes (for most outcomes) is in contrast with high SoE of the effect of CPAP to improve AHI and other sleep and symptom measures.[11, 12] 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. However, until these measures have been validated as appropriate indicators of long-term clinical outcomes, it remains difficult to interpret whether their findings (e.g., of reduced AHI) are of any value to patients. Despite biological plausibility, it is arguably incorrect to simply assume that changes in AHI are associated with, or mediate, improvements in long-term clinical outcomes. Furthermore, most researchers at least implicitly assume that AHI is the "best" intermediate outcome to measure. 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. Until well-conducted studies test the validity of the range of sleep (and symptom) measures to predict important clinical outcomes, it remains difficult to interpret trials of putative intermediate outcomes.

We mostly found low SoE of no effect (or no clinically significant effect) on long-term clinical outcomes of interest. Thus, additional evidence is needed, and it would not be unexpected for the new evidence to change the conclusions regarding all clinical outcomes. The main implication is, thus, that additional studies are needed before we have a clear understanding of the potential effects of CPAP on long-term outcomes for patients with OSA. Currently, clinicians and policymakers can most accurately claim that CPAP may have some clinical benefits for patients with OSA, but that our understanding of the effects of CPAP is likely to change with future evidence. 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 reduce clinical outcomes. From prior comparative effectiveness research conducted for the Agency for Healthcare Research and Quality (AHRQ),[12] it is well-understood that CPAP is effective to improve (reduce) AHI (when used)—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 clinical outcomes.

Future Research

Several types of new studies are needed before stronger evidence is available to allow better clinical and policy decisionmaking.

New, large RCTs are needed. The largest trial to date, SAVE, was still underpowered to provide precise estimates of effects of CPAP treatment on most clinical outcomes. While there is a good research rationale for powering studies for a relatively large effect on a composite outcome, as was done by SAVE and other trials, such an approach limits the value of the studies for patients and clinicians (who generally don't think in terms of arbitrary combinations of clinical outcomes). Although a generic recommendation for future research, it is the case that we need large trials in which patients are followed for a long enough period of time so that both statistically significant and clinically significant differences in specific important clinical outcomes are discernable. To allow for improved decisionmaking by patients, clinicians, and policymakers, additional evidence is needed for all long-term clinical outcomes of interest to patients that are addressed in this report. Better standardization of outcomes (particularly

"composite cardiovascular 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 found at least five different definitions for major adverse cardiovascular event (MACE).[160] Notably, none of the major trials in their review used the same definition. In a Delphi process, the highest-rated definition included cardiac death, myocardial infarction, and nonfatal cardiac arrest[161] (notably, omitting revascularization, unstable angina, heart failure, cerebrovascular events, and arrhythmias, which were all commonly included in CPAP study composite CV outcomes). Similarly, the European Medicines Agency, in their 2016 assessment of cardiovascular 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).[162]

Where feasible, it is preferable to evaluate incident (or resolved) diagnostic conditions, such as depression rather than psychometric tests of severity of symptoms. This would, again, improve interpretation of findings and usefulness for decisionmaking. Where psychometric measures are most important, such as for evaluation of quality of life, these should be validated in patients with OSA to ensure that changes in the measures are clinically meaningful and interpretable. Ideally, evaluated quality of life measures should be demonstrated to be health-related.

Furthermore, ideally, future studies will implement eligibility criteria that are broadly applicable to patients being considered for CPAP use. This implies including patients with a wide range in measures of OSA "severity", ages, and comorbidities. Also, ideally, studies will be large enough (and generalizable enough) to allow meaningful comparisons of subgroups of patients based on such factors as different definitions of OSA (e.g., different diagnostic parameter thresholds with or without comorbidities or symptoms), different severity thresholds, different comorbidities (e.g., obesity or glucose intolerance, which may be confounders affecting both risk of OSA and effectiveness of CPAP), men versus women, and different age categories. The goal should be a determination of which patients would most benefit from CPAP treatment. Such a trials will require large resource commitments.

Under the assumption that future evidence supports that CPAP benefits long-term outcomes, evidence is needed to address the reasons for noncompliance and address whether these issues can be mitigated. Additional evidence will also be needed to determine the effect that varying levels of compliance (e.g., by hours per night or nights per week of use) has on effectiveness of CPAP prescription. Of interest would be whether there is a continuous (e.g., linear) association between level of compliance and outcomes or whether there is a threshold response (e.g., ≥4 hours per night on 70% of nights). Further understanding of the role of compliance in CPAP effectiveness would be of particular importance if future studies support clinical benefits of CPAP, but continue to fail to find differences between effects based on CPAP compliance or use.

Future trials should continue to minimize risk of bias and should be well-reported using consistent methodologies. While, ideally, patients and clinicians should be blinded, we believe it is reasonable to consider the use of sham CPAP to be impractical for the number of study participants and for the duration of treatment required for adequately powered trials. Nevertheless, it is very important to minimize dropouts (loss to followup) and crossovers. As discussed, one of the largest concerns regarding the overall body of evidence relates to a lack of consistency across studies (and sometimes within studies) of definitions and criteria used to assess polysomnography measures and to diagnose OSA. It is critical that both standard criteria are applied and that these criteria are applied consistently and accurately across patients. Fully

transparent reporting about what polysomnography devices were used, where sleep studies were done, how sleep and breathing measures were measured, and what thresholds and other criteria were applied to each measure. The reader of the study should be able to accurately reproduce how study eligibility criteria (including such factors as AHI) were implemented.

Formal assessments of the validity of putative intermediate and surrogate outcomes are needed. 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 clinical outcomes, we found no studies that have tested this hypothesis. Ideally, formal assessments of validity need to be conducted, including mediation analyses that test the indirect effects as represented in clinically meaningful directed acyclic graphs (e.g., see Figure Corr.03). Simpler within-study correlations between changes in the putative intermediate/surrogate measures and effects on long-term clinical outcomes would also provide supporting evidence.

In addition to new studies, where feasible, existing studies should reexamine unreported and/or unanalyzed data to address outcomes of interest and correlations between AHI (and other putative intermediate or surrogate measures) and outcomes. Individual studies should be reanalyzed and/or individual participant-level meta-analysis should be conducted to more accurately combine study results and to allow investigation of the heterogeneity of treatment effect (subgroup differences) and the impact of different polysomnography measure definitions on treatment effects. Such secondary analyses could identify specific areas of study for future, potentially definitive, RCTs.

Existing or new studies should also be analyzed to evaluate the validity of changes in a host of potential intermediate and surrogate measures as potential surrogate or mediator measures of clinical outcomes. Although AHI and ODI have become commonly-evaluated measures to assess OSA severity (and to diagnose OSA), it is important to assess the relative validity of AHI as a surrogate (or mediator) measure against other measures such as apnea index, respiratory effort related awakenings, and also ESS and other symptom measures.

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 polysomnography criteria. Reporting of specific diagnosis and polysomnography criteria is generally poor.

RCTs have not found significant effects of CPAP on long-term clinical outcomes for adults with OSA. There is low strength of evidence from RCTs that CPAP has no effect on all-cause mortality, stroke, acute myocardial infarction, composite cardiovascular outcomes, accidents, incident diabetes, and on clinically significant improvements in (unvalidated) measures of depression and anxiety symptoms, cognitive function, or quality of life. The effect of CPAP on other long-term clinical outcomes is unclear due to insufficient evidence related to sparse studies and/or highly imprecise estimates. Inclusion of evidence from adjusted NRCSs mostly does not alter these conclusions, except that with NRCS evidence, there is low strength of evidence that CPAP reduces the risk of all-cause death, possibly particularly among older adults and others at relatively high risk of dying. The NRCS evidence does not alter conclusions for other outcomes. Given the low strength of evidence across outcomes, additional studies may substantially alter these conclusions.

Studies comparing CPAP with mandibular advancement devices provided low strength of evidence 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 moderate strength of evidence (based on two RCTs) that functional status remains similar between patients prescribed either fixed CPAP or autoCPAP. Other long-term clinical 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 clinical outcomes. Across studies, there was insufficient evidence to support whether changes in intermediate or surrogate measures are correlated with clinical 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 clinical 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, www.aasmnet.org, American Academy of Sleep Medicine, Darien, IL. 2020.

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

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

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

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

21. 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.

22. ECRI Guidelines Trust. https://guidelines.ecri.org. 2019.

23. U.S. Food and Drug Administration. Medical Device Databases. https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/medical-device-databases. 2020.

24. 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.

25. 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.

26. 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.

27. 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.

28. Burzykowski T, Molenberghs G, Buyse M. The evaluation of surrogate endpoints: Springer Science & Business Media; 2006.

29. 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.

30. 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.

31. 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.

32. 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.

33. Shahar E. Apnea-hypopnea index: time to wake up. Nat Sci Sleep. 2014;6:51-6. doi: 10.2147/nss.S61853. PMID: 24741336.

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

39. 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.

40. 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.

41. 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.

42. Berry RB, Brooks R, Gamaldo CE, et al. The AASM manual for the scoring of sleep and associated events. Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine. 2012;176:2012.

43. 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.

44. 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.

45. 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.

46. 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.

47. 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.

48. 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.

49. 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.

50. 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.

51. 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.

52. 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.

53. 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.

54. 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.

55. 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.

56. 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.

57. 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.

58. 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.

59. 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.

60. 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.

61. 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.

62. 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.

63. 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.

64. 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.

65. 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.

66. 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.

67. 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.

68. 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.

69. 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.

70. 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.

71. 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.

72. 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.

73. 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.

74. 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.

75. 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.

76. 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.

77. US 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.

78. US Food and Drug Administration. Devices@FDA. https://www.accessdata.fda.gov/scripts/cdrh/devicesatfda/. Accessed Mar. 1, 2020.

79. 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.

80. 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.

81. 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.

82. 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.

83. 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.

84. 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.

85. 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.

86. 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.

87. 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.

88. 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.

89. 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.

90. 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.

91. 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.

92. 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.

93. 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.

94. 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.

95. 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.

96. 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.

97. 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.

98. 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.

99. de Batlle JB, S; Turino, C; Escarrabill, J; Sánchez-de-la-Torre, M; Woehrle, H; Barbé, F. . Continuous Positive Airway Pressure (CPAP) Treatment Reduces Mortality at the Population Level. American Journal of Respiratory and Critical Care Medicine. 2018;197:A4399.

100. Jara SM PA, Maynard C, Weaver EM. . Long-term survival in veterans with sleep apnea. SLEEP. 2018;41.

101. Nakamura HaK, T. and Takara, C. and Tsukayama, A. and Tohyama, K. and Matsumoto, T. and Iseki, K.. A retrospective analysis of 4000 patients with obstructive sleep apnea in Okinawa, Japan. Sleep and Biological Rhythms. 2009;7(2).

102. Nakamura KaN, H. and Iseki, C. and Toyama, K. and Iseki, K. Survival benefit of continuous positive airway pressure in patients with obstructive sleep apnea: A propensity-score matched analysis. Sleep. 2019;42.

103. 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.

104. 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.

105. 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.

106. 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.

107. 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.

108. 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.

109. 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.

110. 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.

111. 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.

112. 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.

113. 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.

114. 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.

115. 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.

116. 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.

117. 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.

118. 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.

119. 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.

120. 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.

121. 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.

122. 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.

123. 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.

124. 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.

125. Ricker JH, Axelrod BN, Houtler BD. Clinical validation of the oral trail making test. Neuropsychiatry, Neuropsychology, & Behavioral Neurology. 1996.

126. 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.

127. 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.

128. 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.

129. 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.

130. 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.

131. 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.

132. 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.

133. 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.

134. 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.

135. 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.

136. 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.

137. 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.

138. 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.

139. 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.

140. 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.

141. 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.

142. 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.

143. Russell A BJ, Spallek M. ALSWH Data Dictionary Supplement Section 2 Core Survey Dataset 2.3 Health-related Quality of Life Variables SF-36 – Standardised Component Scores. 2009. http://www.alswh.org.au/InfoData/DataDictionarySupplement/DDSSection2SF36.pdf/.

144. 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.

145. 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.

146. 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.

147. 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.

148. 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.

149. 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.

150. 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.

151. 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.

152. 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.

153. 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.

154. VanderWeele T. Explanation in causal inference: methods for mediation and interaction: Oxford University Press; 2015.

155. 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.

156. 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.

157. 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.

158. 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.

159. 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.

160. 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.

161. 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.

162. 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.

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
BiPAP	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