Screening for Obstructive Sleep Apnea in Adults
Evidence Report and Systematic Review
for the US Preventive Services Task Force

Daniel E. Jonas, MD, MPH; Halle R. Amick, MSPH; Cynthia Feltner, MD, MPH; Rachel Palmieri Weber, PhD; Marina Arvanitis, MD, MPH; Alexander Stine, BA; Linda Lux, MPA; Russell P. Harris, MD, MPH

IMPORTANCE Many adverse health outcomes are associated with obstructive sleep apnea (OSA).

OBJECTIVE To review primary care–relevant evidence on screening adults for OSA, test accuracy, and treatment of OSA, to inform the US Preventive Services Task Force.

DATA SOURCES MEDLINE, Cochrane Library, EMBASE, and trial registries through October 2015, references, and experts, with surveillance of the literature through October 5, 2016.

STUDY SELECTION English-language randomized clinical trials (RCTs); studies evaluating accuracy of screening questionnaires or prediction tools, diagnostic accuracy of portable monitors, or association between apnea-hypopnea index (AHI) and health outcomes among community-based participants.

DATA EXTRACTION AND SYNTHESIS Two investigators independently reviewed abstracts and full-text articles. When multiple similar studies were available, random-effects meta-analyses were conducted.

MAIN OUTCOMES AND MEASURES Sensitivity, specificity, area under the curve (AUC), AHI, Epworth Sleepiness Scale (ESS) scores, blood pressure, mortality, cardiovascular events, motor vehicle crashes, quality of life, and harms.

RESULTS A total of 110 studies were included (N = 46 188). No RCTs compared screening with no screening. In 2 studies (n = 702), the screening accuracy of the multivariable apnea prediction score followed by home portable monitor testing for detecting severe OSA syndrome (AHI ≥30 and ESS score >10) was AUC 0.80 (95% CI, 0.78 to 0.82) and 0.83 (95% CI, 0.77 to 0.90), respectively, but the studies oversampled high-risk participants and those with OSA and OSA syndrome. No studies prospectively evaluated screening tools to report calibration or clinical utility for improving health outcomes. Meta-analysis found that continuous positive airway pressure (CPAP) compared with sham was significantly associated with reduction of AHI (weighted mean difference [WMD], −33.8 [95% CI, −42.0 to −25.6]; 13 trials, 543 participants), excessive sleepiness assessed by ESS score (WMD, −2.0 [95% CI, −2.6 to −1.4]; 22 trials, 2721 participants), diurnal systolic blood pressure (WMD, −2.4 points [95% CI, −3.9 to −0.9]; 15 trials, 1190 participants), and diurnal diastolic blood pressure (WMD, −1.3 points [95% CI, −2.2 to −0.4]; 15 trials, 1190 participants). CPAP was associated with modest improvement in sleep-related quality of life (Cohen d, 0.28 [95% CI, 0.14 to 0.42]; 13 trials, 2325 participants). Mandibular advancement devices (MADs) and weight loss programs were also associated with reduced AHI and excessive sleepiness. Common adverse effects of CPAP and MADs included oral or nasal dryness, irritation, and pain, among others. In cohort studies, there was a consistent association between AHI and all-cause mortality.

CONCLUSIONS AND RELEVANCE There is uncertainty about the accuracy or clinical utility of all potential screening tools. Multiple treatments for OSA reduce AHI, ESS scores, and blood pressure. Trials of CPAP and other treatments have not established whether treatment reduces mortality or improves most other health outcomes, except for modest improvement in sleep-related quality of life.


Author Affiliations: Department of Medicine, University of North Carolina at Chapel Hill (Jonas, Feltner, Arvanitis, Harris); RTI International–University of North Carolina at Chapel Hill Evidence-based Practice Center (Jonas, Amick, Feltner, Weber, Stine, Lux, Harris); Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill (Jonas, Amick, Feltner, Weber, Arvanitis, Harris); Now with the Department of Medicine, Northwestern University, Chicago, Illinois (Arvanitis); RTI International, Research Triangle Park, North Carolina (Stine, Lux); Now with the Center for Cognitive Neuroscience, Duke University, Durham, North Carolina (Stine).

Corresponding Author: Daniel E. Jonas, MD, MPH, 5034 Old Clinic Bldg, Chapel Hill, NC 27599 (daniel.jonas@med.unc.edu).
Obstructive sleep apnea (OSA) (Table 1) has been associated with an increased risk of many adverse health outcomes, including motor vehicle crashes, cognitive impairment, cardiovascular events, atrial fibrillation, stroke, and mortality. However, there is controversy in the literature regarding the extent to which OSA independently contributes to various outcomes beyond the contributions of age, body mass index (BMI), and other potential confounders. OSA is common, with prevalence around 15% in men and 5% in women (ages 30-70 years), based on either an apnea-hypopnea index (AHI) of 15 or greater or an AHI of 5 or greater plus symptoms of disturbed sleep.

Screening to identify unrecognized OSA followed by appropriate treatment might improve sleep quality and normalize the AHI and oxygen saturation levels to prevent adverse health outcomes. Potential screening strategies include questionnaires and clinical prediction tools that comprise combinations of subjective and objective findings. For people who screen positive, diagnostic polysomnography in a sleep facility or home-based testing with a portable monitor could be used to determine whether they have OSA.

To inform a recommendation by the US Preventive Services Task Force (USPSTF), the evidence on test accuracy and benefits and harms of screening and treatment for OSA in populations and settings relevant to US primary care was reviewed.

Methods

Scope of Review
Detailed methods are available in the full evidence report at https://www.uspreventiveservicestaskforce.org/Page/Document/ final-evidence-review152/obstructive-sleep-apnea-in-adults-screening. Additional subgroup analyses (by OSA severity, baseline sleepiness, and baseline blood pressure) and sensitivity analyses conducted to explore heterogeneity or robustness of findings are available in the full evidence report. Figure 1 shows the analytic framework and key questions that guided the review.

Data Sources and Searches
We searched PubMed/MEDLINE, the Cochrane Library, and EMBASE for English-language articles published through October 2015. ClinicalTrials.gov and the World Health Organization International Clinical Trials Registry Platform were also searched for unpublished literature. The search strategies for PubMed and Cochrane databases are detailed in the eMethods in the Supplement. To supplement electronic searches, the reference lists of pertinent articles were reviewed, as well as all studies suggested by reviewers or comments received during public commenting periods. Since October 2015, we conducted ongoing surveillance through article alerts and targeted searches of high-impact journals to identify major studies published in the interim that may affect the conclusions or understanding of the evidence and therefore the related USPSTF recommendation. The last surveillance was conducted on October 5, 2016.

Study Selection
Two investigators independently reviewed titles, abstracts, and full-text articles to determine eligibility using prespecified criteria for each key question (KQ) (eTable 1 in the Supplement). Disagreements were resolved by discussion. The review included English-language studies of adults conducted in countries categorized as “very high” on the Human Development Index. Only studies rated as good or fair quality using predefined criteria and definitions developed by the USPSTF and adapted for this topic (eTable 2 in the Supplement) were included. The review excluded studies of people with acute conditions (eg, stroke) that can trigger onset of OSA and studies focused on screening, diagnosis, or treatment of OSA among persons with rare conditions (eg, acromegaly) for whom testing for OSA would be considered part of management for their disease.

For the overarching question regarding direct evidence that screening improves health outcomes (KQ1) and the question on accuracy of clinical prediction tools or screening questionnaires (KQ2), studies were required to enroll asymptomatic adults or persons with unrecognized symptoms of OSA; referral populations were not eligible. For KQ1, randomized clinical trials (RCTs) comparing screened with nonscreened groups were eligible. For KQ2, studies that evaluated screening questionnaires or clinical prediction tools (alone or followed by home-based portable monitoring) compared with overnight polysomnography conducted in a sleep laboratory were eligible. Studies of people referred to sleep laboratories because of concern for OSA were excluded, and studies in which only a subgroup (usually the highest-risk group) underwent polysomnography were excluded because of concern for verification bias. Clinical prediction tools were required to include multiple factors.

For diagnostic test accuracy (KQ3) and harms associated with screening and diagnostic tests (KQ7), referral populations were also eligible (in addition to the populations eligible for KQ1 and KQ2). For KQ3, good-quality, recent systematic reviews comparing portable monitors (Table 2 describes the types of monitors) with polysomnography conducted in a sleep laboratory were eligible. Multiple good-quality, recent, and relevant systematic reviews for KQ3 were identified; primary studies published after the search cutoffs of the most recent systematic reviews were also included. For KQ7, studies eligible for KQ1, KQ2, or KQ3 that reported false-positive results leading to unnecessary treatment, anxiety, condition-specific distress, or stigma were eligible.

For benefits and harms of treatment (KQ4, KQ5, and KQ8), RCTs enrolling people with a confirmed diagnosis of OSA were eligible; studies could include asymptomatic adults, symptomatic adults, or both. Studies evaluating continuous positive airway pressure (CPAP), mandibular advancement devices (MADs), surgery, and weight loss programs were included; other treatments were not eligible (eg, oropharyngeal exercises). For KQ8, prospective cohort studies with at least 100 participants that reported harms of surgical interventions were also eligible.

For the association between AHI and health outcomes (KQ6), prospective cohort studies that followed up participants for at least 1 year were included. Studies were excluded that focused primarily on central sleep apnea, enrolled patients hospitalized for acute events, enrolled patients in a perioperative period, or did not address potential confounding.

Data Extraction and Quality Assessment
For each included study, one investigator extracted information about the populations, tests or treatments, comparators, outcomes,
settings, and designs, and a second investigator reviewed for completeness and accuracy. Two independent investigators assessed the quality of studies as good, fair, or poor. Disagreements were resolved by discussion.

Data Synthesis and Analysis
Findings for each question were summarized in tabular and narrative form. To determine whether meta-analyses were appropriate, the clinical and methodological heterogeneity of the studies was assessed following established guidance.22 When multiple similar studies were available, quantitative synthesis was conducted with random-effects models using the inverse-variance weighted method (DerSimonian and Laird) to estimate pooled effects.23 For all quantitative syntheses, the $I^2$ statistic was calculated to assess statistical heterogeneity in effects between studies.24,25 Quantitative analyses were conducted using Comprehensive Meta-Analysis version 3.3 (Biostat Inc) and Stata version 14 (StataCorp). Statistical significance was assumed when 95% CIs of pooled results did not cross the null (ie, 0 or 1, depending on the effect measure). All testing was 2-sided. This review covered a wide range of outcome measures and instruments; key measures and questionnaires are summarized in eTable 3 in the Supplement.

For KQ4 and KQ5 the weighted mean difference (WMD) between intervention and control was calculated for continuous outcomes; when multiple scales were combined in a single meta-analysis (for sleep-related quality of life), we used the standardized mean difference, Cohen d. For Cohen d, a value of 0.20 is often interpreted as a small effect size, 0.50 as a medium effect size, and 0.80 as a large effect size.26 For meta-analyses of CPAP and MAD treatments, pooled estimates were calculated separately for studies using sham controls and those using other controls. Parallel trials and crossover trials were combined, but subgroup analyses were conducted to explore whether findings differed by this design feature.

For KQ6, we conducted meta-analyses of adjusted hazard ratios (HRs) and 95% confidence intervals for all-cause mortality. The HRs were converted to a log scale, and standard errors of the log HRs were calculated to normalize distributions and stabilize variances. The metan command with the eform command in Stata was then used to estimate pooled HRs. Analyses were by AHI thresholds corresponding to OSA severity categories.

Results
A total of 110 studies (127 articles) with $N = 46188$ participants were included (Figure 2). Individual study quality ratings are reported in eTables 4 through 12 in the Supplement. The main results for each key question are summarized below; additional details and analyses are available in the full evidence report.27

Benefits of Screening
Key Question 1a. Does screening for OSA in adults improve health outcomes?
Key Question 1b. Does the evidence differ for subgroups defined by age, sex, BMI, or OSA severity?

No eligible studies were identified.

Table I. Obstructive Sleep Apnea–Related Terms and Definitions

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apnea</td>
<td>Cessation of airflow for at least 10 s$^1,2$</td>
</tr>
</tbody>
</table>
| Hypopnea              | Reduction in airflow by at least 30% for at least 10 s with decrease in oxygen saturation | $^3$
| AHI$^4,5$             | Number of apneas and hypopneas per h of sleep                            |
| OSA$^4$               | AHI ≥5 to <15                                                            |
| Moderate$^1,3$        | AHI ≥15 to <30                                                           |
| Severe$^1,3$          | AHI ≥30                                                                  |
| Obstructive sleep apnea syndrome | AHI ≥5 with evidence of daytime sleepiness$^3,4,5$                      |

Abbreviations: AHI, apnea-hypopnea index; OSA, obstructive sleep apnea.

4 OSA occurs when airflow is absent or substantially reduced because of upper airway obstruction but breathing effort persists. OSA severity is usually categorized using the AHI as assessed by a sleep study (polysomnography).

5 The respiratory disturbance index is a measure similar to the AHI, but it also includes the number of respiratory effort–related arousals per hour of sleep (in addition to apnea events and hypopnea events).

6 The AHI incorporates both obstructive and central apnea and hypopnea events, and significantly elevated AHI is not synonymous with OSA (because it can indicate OSA, central sleep apnea, or mixed sleep apnea—both OSA and central sleep apnea).

7 Both the Centers for Medicare & Medicaid Services and the American Academy of Sleep Medicine define OSA as an AHI or respiratory disturbance index of at least 15 events per hour, or at least 5 events per hour with documented symptoms (eg, excessive daytime sleepiness, impaired cognition, mood disorders, or insomnia; waking up breath-holding, gasping, or choking; or documented hypertension, ischemic heart disease, or history of stroke).4,6

Accuracy of Clinical Prediction Tools or Screening Questionnaires
Key Question 2a. What is the accuracy of currently existing clinical prediction tools or screening questionnaires in identifying persons in the general population who are more or less likely to have OSA?

Key Question 2b. What is the accuracy of multistep screening approaches, such as using a questionnaire or prediction tool followed by overnight home-based testing, in identifying persons in the general population who are more or less likely to have OSA?

Three studies were included (Table 3)28-30 One evaluated the Berlin Questionnaire,28 and 2 evaluated the Multivariable Apnea Prediction (MVAP) score, alone and when followed by in-home portable monitoring.29,30 Details of the questions and scoring are reported in the eBackground in the Supplement.

The study evaluating the Berlin Questionnaire randomly sampled Norwegians from the National Population Register (55% response rate: 16302/29258).28 Of those completing the questionnaire, 24% were classified as high risk and 518 had undergone in-hospital polysomnography. Of those 518, mean age was 48 years, 45% were female, mean BMI was 28 (calculated as weight in kilograms divided by height in meters squared), and median AHI was 6.4. Although the group undergoing polysomnography oversampled high-risk participants (70% were high risk), the analyses adjusted for bias in the sampling to report estimated screening properties for the general population. The study found suboptimal screening properties (for AHI ≥5: sensitivity, 37.2%; specificity, 84%; for AHI ≥15: sensitivity, 43%; specificity, 79.7%)
The unadjusted analyses showed much better sensitivity but worse specificity (for AHI ≥5: sensitivity, 79.4%; specificity, 40.5%; for AHI ≥15: sensitivity, 82.8%; specificity, 34.9%), likely reflecting spectrum bias.

Both studies assessing the MVAP included highly selected patients.29,30 One study evaluated Medicare recipients (n = 452), most (74%) of whom had daytime sleepiness.29 The percentage with OSA was not reported, but 27% had obstructive sleep apnea syndrome (OSAS), defined as AHI ≥5 or greater and Epworth Sleepiness Scale (ESS) score greater than 10. The other study evaluated patients with hypertension (n = 250).30 Eighty percent of participants had OSA (AHI ≥5); of those, 22% had moderate and 25% had severe OSA; 25% of all participants had OSAS. Mean ages of participants were 7129 and 5330 years; 60% to 64% were non-white; and mean BMIs were 30 to 32, respectively. Key quality limitations included concern for attrition bias30 and moderate concern for selection bias or spectrum bias (with high prevalence of OSA, OSAS, and/or daytime sleepiness among those undergoing polysomnography).29,30

Both studies reported operating characteristics of MVAP to predict severe OSAS (AHI ≥30 and ESS score >10) (Table 4).
Briefly (area under the curve [AUC], 0.78 [95% CI, 0.71 to 0.85]), whereas the other study found inadequate discrimination (AUC, 0.68 [95% CI, 0.67 to 0.70]). An AUC less than 0.70 has been considered to indicate inadequate discrimination. Both studies also reported measures of discrimination for the MVAP score followed by in-home portable monitoring (Table 4). The studies by Morales et al and Gurubhagavatula et al reported characteristics to predict severe OSA using different portable monitor–based AHI cut-offs (ie, 15 and 18). Both found better operating characteristics when using MVAP followed by in-home portable monitoring (AUC, 0.80–0.83) than when using MVAP alone. The study of participants with hypertension also reported operating characteristics of MVAP and MVAP followed by in-home portable monitoring to predict any OSA (AHI ≥5 and ESS score >10). It found inadequate discrimination (Table 4).

**Diagnostic Accuracy and Reliability of Portable Monitors**

**Key Question 3a.** What is the accuracy and reliability of diagnostic tests for OSA?

**Key Question 3b.** Do the accuracy and reliability of diagnostic tests for OSA differ for subgroups defined by age, sex, or BMI?

We included 3 studies evaluating type II portable monitors, 1 systematic review and 2 subsequent studies evaluating type III portable monitors, and 1 systematic review and 14 subsequent studies evaluating type IV portable monitors. Study participants were generally those referred to sleep units for suspected sleep apnea. No studies were found that identified participants via screening to provide evidence on asymptomatic patients or those with unrecognized symptoms, although detailed reporting of reasons for referral was generally limited. Details of individual study characteristics and results are provided in the eResults and eTables 13 through 22 in the Supplement.

**Table 5** summarizes the range of sensitivities, specificities, and AUCs by type of portable monitor for AHI thresholds of 5, 15, and 30. The best evidence comes from systematic reviews that reported sensitivities of 93% (pooled estimate from in-home studies) and 96% (pooled estimate from in-laboratory studies) for type III portable monitors and at least 85% for type IV portable monitors. Type II monitors usually measure the same channels as type I monitors but are portable.

Corresponding specificities were 60% and 76% for in-home and in-laboratory type III portable monitors, respectively, and ranged from 50% to 100% for type IV portable monitors. Sensitivities decreased and specificities increased for detecting moderate or greater OSA (AHI ≥15) or severe OSA (AHI ≥30). The ranges of sensitivity and specificity reported across studies for type IV monitors were wide.

### Benefits of Treatment

**Key Question 4a.** How much does treatment with CPAP, MADs, surgery, or weight loss programs improve intermediate outcomes (AHI, blood pressure, or daytime sleepiness) in persons with OSA?

**Key Question 4b.** Do the benefits of treatment (for intermediate outcomes) differ for subgroups defined by age, sex, BMI, or OSA severity?

Included were 76 RCTs: 56 trials evaluated CPAP (eTables 23 and 24 in the Supplement), 53-112 10 trials evaluated MADs (eTable 25 in the Supplement), 98,105,113-122 6 trials evaluated surgical interventions (eTable 26 in the Supplement), 123-128 and 6 trials evaluated weight loss, diet, and exercise programs (eTable 27 in the Supplement). None of the trials focused on participants who were screen-detected in primary care settings.

#### Continuous Positive Airway Pressure

Most studies identified participants from sleep clinics or referrals. Duration of treatment ranged from 1 week to 4 years. Most trials lasted for 12 weeks or less, but 5 trials treated participants for 24 weeks or longer. Including 2 that followed up participants for 52 weeks and 1 that did so for a median of 4 years. Mean age was 40s to 50s in most studies (range, 42-71). The majority of participants in most trials were men, with 44 trials reporting that less than one-third of participants were women. Mean BMI was 30 to 35 in most trials (range, 27-39). Mean or median baseline AHI (or similar measure) was in the severe OSA range (AHI ≥30) for more than 75% of trials; 8 trials reported it in the moderate OSA range (15-30) and 4 reported it in the mild OSA range (9-15). Mean baseline ESS score was 10 or more in 33 trials, indicating excessive daytime sleepiness. Ten trials reported...
a mean baseline ESS score less than 10,55,59,63,72,87,96,97,99,103,106 and 13 trials did not report baseline ESS score.

For AHI, trials reporting sufficient data for meta-analysis followed up patients for 12 weeks or less. The meta-analyses found that CPAP was associated with reduction of AHI compared with sham CPAP (WMD, −33.8 [95% CI, −42.0 to −25.6]; 13 trials, 543 participants) and other controls (WMD, −25.8 [95% CI, −34.2 to −17.5]; 6 trials, 294 participants) (eFigures 1 and 2 in the Supplement). All individual studies reported end-point AHI values of 10 or less for CPAP-treated groups, and most were normal (<5).

Thirty-four trials reported sufficient ESS data to include in meta-analyses. Most were 12 weeks or less in duration; 5 followed up participants for 24 weeks,70,99 48 to 52 weeks,96,107 or longer.97 The meta-analyses found that CPAP was associated with reduction of ESS scores compared with sham CPAP (WMD, −2.0 [95% CI, −2.6 to −1.4]; 22 trials, 2721 participants) and other controls (WMD, −2.2 [95% CI, −2.8 to −1.6]; 12 trials, 2488 participants) (eFigures 3 and 4 in the Supplement). Among the 27 trials with mean or median baseline ESS scores of 10 or greater (mean baseline ESS score was 12.7 among them) or those that provided subgroup analyses for the participants with excessive sleepiness, the subgroup analysis found a similar result (WMD, −2.4 [95% CI, −2.9 to −1.9]) (eFigure 5 in the Supplement). Twenty-three of those 27 trials reported mean end-point ESS scores less than 10 for the CPAP group (mean end-point ESS score for the 23 trials was <8).

Twenty-nine trials reported sufficient blood pressure data to include in meta-analyses. Blood pressure outcomes were reported in a variety of ways; most commonly, diurnal systolic and diurnal diastolic blood pressure. Most trials were 12 weeks or less in duration; 3 followed up participants for 24 to 52 weeks.96,99,107 The meta-analyses found that CPAP was associated with reduction of diurnal systolic blood pressure by 2 to 3 points (WMD, −2.4 [95% CI, −3.9 to −0.9]; 15 trials, 1190 participants) (eFigure 6 in the Supplement) and diurnal diastolic blood pressure by more than 1 point (WMD, −1.3 [95% CI, −2.2 to −0.4]; 15 trials, 1190 participants) (eFigure 7 in the Supplement) compared with sham CPAP. Reduction in 24-hour mean arterial pressure was about 2 points with...
CPAP compared with sham CPAP (WMD, −2.1 [95% CI, −3.2 to −1.0]); 5 trials, 621 participants) (eFigure 8 in the Supplement). Among the 6 studies that provided results for participants with uncontrolled hypertension, the subgroup analysis found similar but slightly larger effect sizes (eFigures 9 and 10 in the Supplement); reductions of −2.5 points for diurnal systolic blood pressure, −2.1 points for diurnal diastolic blood pressure, and −2.7 points for 24-hour mean arterial pressure.

Mandibular Advancement Devices
Six of the 10 included RCTs compared MADs with sham devices. Comparators used in other RCTs were a placebo tablet, no treatment, and conservative management with weight loss. All studies recruited participants with known or suspected OSA from specialty clinics. Treatment durations ranged from 4 to 12 weeks for most studies, but 1 lasted only 1 week and 1 lasted 24 weeks. Mean age of participants ranged from 45 to 59 years. The majority of participants in all trials were men, with women comprising 17% to 25% of participants in the 9 trials reporting sex. All studies included participants with mild to moderate OSA, and 6 also included participants with severe OSA. Mean baseline ESS scores ranged from 11 to 14.

The meta-analyses found that MADs were associated with greater improvement in AH1 than sham devices (−12.6 [95% CI, −36.2 to −16.6] and −10.5 [95% CI, −16.9 to −4.1], respectively). Further details of the characteristics and results of trials that evaluated surgical interventions are provided in the eResults and eFigures 21 and 22 in the Supplement.

Airway Surgery and Bariatric Surgery
Six trials each evaluated a different surgical technique, including radiofrequency surgery of the soft palate, temperature-controlled radiofrequency tissue ablation, uvulopalatopharyngoplasty, laser-assisted uvulopalatoplasty, septoplasty, and bariatric surgery (eTable 26 in the Supplement). Sample sizes ranged from 32 to 67. Overall, the trials provided limited evidence and found no significant reduction in AH1, ESS scores, or blood pressure, with the exception of the trials of uvulopalatopharyngoplasty and laser-assisted uvulopalatoplasty, which found greater reductions in AH1 for surgery than for no treatment (−26.4 [95% CI, −36.2 to −16.6] and −10.5 [95% CI, −16.9 to −4.1], respectively). Further details of the characteristics and results of trials that evaluated surgical interventions are provided in the eResults and eFigures 21 and 22 in the Supplement.

Weight Loss, Diet, and Exercise Interventions
Six trials evaluated weight loss programs (eTable 27 in the Supplement). Each trial evaluated a different intervention
and control—2 interventions focused primarily on exercise,129,133 2 focused primarily on diet,132,136 and 2 used multicomponent lifestyle interventions (exercise, diet, and psychoeducation).130,135 Sample sizes ranged from 26129 to 264.130 Participants were generally identified from sleep clinics, referrals, and advertisements. Duration of follow-up was 4 to 26 weeks for 4 of the trials; the other 2 trials followed up participants for 4 or 5 years.130,136 Mean age ranged from 47 to 61 years. Mean BMI ranged from 30 to 40. Mean AH1 was in the moderate to severe OSA range for 4 of the trials, in the mild range for 1 trial,136 and was moderate to severe but controlled with CPAP use in 1 trial.135 Mean baseline ESS score was 10 or more in 2 trials129,136,138 and less than 10 in 3 trials,132,133,135 and not reported for 1 trial.130 The weight loss achieved by intervention groups was very limited in 1 trial (−0.3 kg),133 modest in another (−2.3 kg),135 and larger in the rest (−5 kg to −20 kg).130,132,138

Four of the 5 trials129,130,132,133,138 reporting AH1 found statistically significant reductions, ranging from −5.8 (95% CI, −9.7 to −1.9) to −23 (−30.1 to −15.9) (efigure 23 in the Supplement). The trial reporting the largest reduction in AH1 (a reduction nearing that achieved by CPAP) also reported a much larger weight reduction than other trials (−20 kg over 9 weeks from a very low energy diet).132 The meta-analysis for AH1 found a WMD of −12.4 (95% CI, −19.4 to −5.5). Three of the 4 trials129,132,133,138 reporting ESS scores found statistically significant reductions, ranging from −3 to −7. The meta-analysis found that weight loss interventions were associated with improvement in ESS scores compared with controls (−3.4 [95% CI, −5.9 to −1.0]; 4 trials, 213 participants) (eFigure 24 in the Supplement). Three trials reported blood pressure outcomes and found no significant differences between treatment and control groups.134–136

Benefits of Treatment

Key Question 5a. Does treatment with CPAP, MADs, surgery, or weight-loss programs improve health outcomes in persons with OSA?

Included were 50 RCTs (eTables 25 through 31 in the Supplement) that reported at least 1 eligible health outcome (47 of these were included in KQ4). Most were short-term RCTs (12 weeks or less) that reported zero or few deaths. None focused on screen-detected patients from primary care settings. The main findings

Table 4. Results of Included Studies: Accuracy of Screening Questionnaires and Clinical Prediction Tools (Key Question 2)

<table>
<thead>
<tr>
<th>Source</th>
<th>Questionnaire/Tool Name and Cutoff Value</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUROC</th>
<th>Calibration</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hrubos-Strom et al,132,135 2011</td>
<td>BQ to predict AH1 ≥5 Cutoff = BQ high risk vs low risk</td>
<td>37.2 (36.0–38.4)</td>
<td>84.0 (83.2–84.7)</td>
<td>NR</td>
<td>NR</td>
<td>PPV, 61.3 (95% CI, 59.7–62.9) NPV, 66.2 (95% CI, 65.3–67.1) LR+, 2.3 (95% CI, 2.2–2.5) LR−, 0.8 (95% CI, 0.7–0.8)</td>
</tr>
<tr>
<td>Morales et al,132,136 2012</td>
<td>MVAP to predict severe OSAS (AH1 ≥30 and ESS score &gt;10) Cutoff = uAH1 0.49</td>
<td>90.0 (NR)</td>
<td>64.4 (NR)</td>
<td>0.78 (0.71–0.85)</td>
<td>NR</td>
<td>LR−, 0.141 NPTP, 1.1% (95% CI, NR)</td>
</tr>
<tr>
<td>Gurubhagavatula et al,133,134 2013</td>
<td>MVAP + uAH1 to predict severe OSAS (AH1 ≥30 and ESS score &gt;10) Cutoff = uAH1 15b</td>
<td>91.5 (NR)</td>
<td>43.9 (NR)</td>
<td>0.68 (0.67–0.70)</td>
<td>NR</td>
<td>LR−, 0.190 NPTP, 0.015 (95% CI, NR)</td>
</tr>
<tr>
<td>MVAP to predict any OSAS (AH1 ≥5 and ESS score &gt;10) Cutoff = 0.559</td>
<td>69.4 (NR)</td>
<td>56.5 (NR)</td>
<td>0.61 (NR)</td>
<td>NR</td>
<td>LR−, 0.524 NPTP, 0.148 (95% CI, NR)</td>
<td></td>
</tr>
<tr>
<td>MVAP + uAH1 to predict severe OSAS (AH1 ≥30 and ESS score &gt;10) Cutoff = uAH1 15b</td>
<td>88.2 (NR)</td>
<td>71.6 (NR)</td>
<td>0.80 (0.78–0.82)</td>
<td>NR</td>
<td>LR−, 0.162 NPTP, 0.015 (95% CI, NR)</td>
<td></td>
</tr>
<tr>
<td>MVAP + uAH1 to predict any OSAS (AH1 ≥5 and ESS score &gt;10) Cutoff = uAH1 13.5b</td>
<td>80.5 (NR)</td>
<td>54.0 (NR)</td>
<td>0.67 (NR)</td>
<td>NR</td>
<td>LR−, 0.349 NPTP, 0.104 (95% CI, NR)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: AH1, apnea-hypopnea index; AUROC, area under the receiver operating characteristic curve; BQ, Berlin Questionnaire; ESS, Epworth Sleepiness Scale; LR, likelihood ratio; MVAP, multivariable apnea prediction; NPTP, negative posttest probability; NPV, negative predictive value; NR, not reported; OSAS, obstructive sleep apnea syndrome; PPV, positive predictive value; uAH1, unattended AH1 from home sleep test.

a Estimates were based on a simulated model that adjusted for oversampling of BQ high-risk participants (not just based on findings for the 518 in the clinical sample).

b Two-stage process using MVAP for everyone and then home testing to determine AH1 for those with an intermediate MVAP score.
are summarized below; additional outcomes for which there were limited data are shown in eTables 29 through 31 in the Supplement and are summarized in the full report.

### Continuous Positive Airway Pressure

Thirty-five RCTs compared CPAP with sham CPAP and are summarized in the full report. Most trials followed up participants for 12 weeks or less; 4 trials measured outcomes over 24 weeks or longer.70,97 One trial included participants for 7 weeks121 and another for 24 weeks.114 All studies included people with an average age in the 40s to 50s (range, 42-71 years). Mean BMI was 30 (range, 27-37). Mean or median baseline AHI (or similar measure) was in the severe OSA range (AHI ≥30) for more than half of trials; 9 trials reported it in the moderate OSA range,75,76,80,87,91,93,97,139 and 5 reported it in the mild OSA range.91,99,103,108,141

#### Table 5. Summary of Accuracy of Diagnostic Tests for Obstructive Sleep Apnea, by Portable Monitor Type (Key Question 3)

<table>
<thead>
<tr>
<th>No. and Design of Studies Contributing to Summary</th>
<th>Range, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polysomnography AHI ≥5</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>Portable Monitor Type II</td>
<td>3 studies of diagnostic accuracy (n = 160)11-25</td>
</tr>
<tr>
<td>Portable Monitor Type III</td>
<td>1 systematic review of 19 studies (n = 1507)26</td>
</tr>
<tr>
<td>Portable Monitor Type IV</td>
<td>1 systematic review of 70 studies (n = 6873)2,1</td>
</tr>
</tbody>
</table>

Abbreviations: AHI, apnea-hypopnea index (events/h); AUC, area under the curve; NR, not reported.

* Three included studies of type IV portable monitors did not contribute results for the AHI thresholds in this table.50-52

† The 2011 systematic review did not report the range of AUC values for the 2007 technology assessment and articles newly included in the 2011 review. The AUC values among the studies newly identified since the 2011 review ranged from 59 to 94.

‡ The 2011 systematic review did not report the range of AUC values for the 2007 technology assessment and articles newly included in the 2011 review. The AUC values among the studies newly identified since the 2011 review ranged from 73 to 95.

### Mandibular Advancement Devices

Included were 6 RCTs assessing the effect of MADs on health outcomes (eTable 30 in the Supplement).98,103,104,116,121,122 Treatment durations ranged from 4 to 12 weeks for most studies, while 1 lasted for only 1 week103 and 1 for 24 weeks.116 All studies included baseline 36-Item Short Form Health Survey (SF-36) mental component score (WMD, 1.2 [95% CI, −0.8 to 3.2]; 8 trials, 1039 participants) (eFigure 25 in the Supplement). The meta-analysis found that CPAP was associated with improved SF-36 physical component score compared with sham CPAP over 12 weeks or less (WMD, 2.3 [95% CI, 0.2 to 4.4]; 7 trials, 648 participants) (eFigure 26 in the Supplement).

Thirteen RCTs assessed sleep-related quality of life—6 using the Sleep Apnea Quality of Life Index (SAQLI)97,93,99,105,107,124 and 7 using the Functional Outcomes of Sleep Questionnaire (FOSQ).55,76,79,86,91,98,102 Most reported outcomes at 12 weeks or less; 2 reported outcomes at 24 weeks (or 6 months)99,142 and 1 at 52 weeks.107 The meta-analysis (combining SAQLI and FOSQ scores) found that CPAP was associated with improved sleep-related quality-of-life scores compared with controls (standardized mean difference, 0.28 [95% CI, 0.14 to 0.42]; 13 trials, 2325 participants) (eFigure 27 in the Supplement). The sensitivity analysis including only studies with mean or median baseline ESS scores of 10 or greater found a similar effect size (0.33 [95% CI, 0.17 to 0.50]; 9 trials, 1709 participants) (eFigure 28 in the Supplement).

Eight RCTs reported on the incidence of 1 or more cardiovascular and cerebrovascular events (eTable 29 in the Supplement).63,70,76,79,91,97,103,107 Overall, too few cardiovascular and cerebrovascular events were observed to draw conclusions.
participants with mild to moderate OSA, and 3 also included participants with severe OSA.\textsuperscript{105,116,121}

Among the 4 trials that reported on mortality over 1 to 12 weeks,\textsuperscript{98,116,121,122} 3 reported no deaths in any participants and 1 reported 1 death in the group that received no treatment.\textsuperscript{116} Five included trials reported at least 1 quality-of-life measure.\textsuperscript{98,105,114,116,122} All 5 used the SF-36; 2 also used the SAQLI,\textsuperscript{105,122} and 2 also used the FOSQ.\textsuperscript{98,122} Overall, results were mixed, with some studies finding no significant benefits of MADs for improving quality of life,\textsuperscript{105,114} some reporting possible benefits for some measures or subscales but not others,\textsuperscript{98,116} and some reporting benefits for some overall quality-of-life scores.\textsuperscript{122} Because of inconsistency, imprecision, and heterogeneity of reporting, findings were insufficient to make conclusions about the potential benefits of MADs for improving quality of life.

Airway Surgery and Bariatric Surgery

Although 5 of the 6 RCTs included in KQ4 that evaluated surgical treatments reported some information about at least 1 health outcome, the trials provided limited evidence to determine whether treatments improve health outcomes. The RCT (n = 60) that compared bariatric surgery with a conventional weight loss program in people with severe OSA\textsuperscript{125} reported greater improvement in quality of life measured by the SF-36 physical component score for those randomized to bariatric surgery at 2 years (between-group difference, 9.3 [95% CI, 0.5 to 18.0]); however, there was no significant difference between groups in the change from baseline SF-36 mental component score (between-group difference, −0.3 [95% CI, −5.3 to 4.8]).\textsuperscript{122} Further details on the results of trials that evaluated surgical interventions are provided in the eResults in the Supplement.

Weight Loss, Diet, and Exercise Interventions

Six RCTs evaluated weight loss programs (eTable 27 in the Supplement).\textsuperscript{129-138} Four RCTs (with a total of 45 participants) assessed mortality, 3 reported no deaths in any group over 9 to 208 weeks,\textsuperscript{130,132,133} and 1 reported 1 death at 52 weeks.\textsuperscript{136} Four RCTs assessed quality of life (eTable 31 in the Supplement).\textsuperscript{129,133,135,136} Overall, findings were mixed, and too few studies reported results for the same intervention and comparison using similar outcome measures to draw conclusions.

Association Between Obstructive Sleep Apnea and Health Outcomes

Key Question 6. Is there an association between AHI and health outcomes?

Six studies reported the association with all-cause mortality,\textsuperscript{144,145,147,150} 3 with cardiovascular mortality,\textsuperscript{12,150,151} 2 with cardiovascular events,\textsuperscript{12,148} and 1 each with cancer-related mortality,\textsuperscript{145} stroke,\textsuperscript{145} cognitive decline,\textsuperscript{143} and cognitive impairment or dementia.\textsuperscript{146} Nine of 11 were conducted in the United States. Most studies followed up patients for 8 to 14 years; follow-up ranged from a mean of 3.4 years\textsuperscript{144} to 22 years.\textsuperscript{145} Three studies included only men; half of the studies included between 45% and 56% women. Mean BMI ranged from 26 to 30 in most studies. Participants were generally untreated for OSA, or analyses were run to exclude those who were treated.

Six studies evaluated AHI as a predictor of all-cause mortality.\textsuperscript{144,145,147,150} Sample sizes ranges from 289\textsuperscript{146} to 6,294.\textsuperscript{151} Mean duration of follow-up ranged from 3.4\textsuperscript{144} to 20 years.\textsuperscript{153} Mean age ranged from 48\textsuperscript{150} to 78 years.\textsuperscript{147} In multivariable analyses, all 6 studies reported that participants with severe or moderate to severe OSA at baseline had a higher risk of death. Variables included in the models are detailed in eTable 33 in the Supplement. Briefly, all included age and some medical conditions in the final model; all considered BMI (although it did not remain in the final model in 1 study); most included smoking, sex, race, hypertension or blood pressure, and diabetes. Comparing mortality for patients with severe or moderate to severe OSA vs controls, meta-analysis found an HR of 2.07 [95% CI, 1.48 to 2.91] (Figure 3). Two studies\textsuperscript{150,151} assessed whether moderate (AHI 15 to 30) or mild (AHI 5 to 15) OSA levels are associated with mortality; neither found a statistically significant association (Figure 3).

Two studies reported evidence for subgroups—either by sex and age\textsuperscript{151} or by presence of sleepiness.\textsuperscript{147} The former used the Sleep Heart Health Study data (n = 6294) and reported that the association between an AHI of 30 or greater and mortality was statistically significant for men 70 years or younger (adjusted HR, 2.09 [95% CI, 1.31 to 3.33]) but not for men older than 70 years (HR, 1.27 [95% CI, 0.86 to 1.86]) or for women of any age (HR, 1.40 [95% CI, 0.89 to 2.22]).\textsuperscript{151} The latter found that the association between AHI of 20 or greater and death was limited to those with excessive daytime sleepiness (determined by self-report of having a problem with feeling sleepy or struggling to stay awake during the daytime more than 3 or 4 times a week) but was not significant for those without excessive daytime sleepiness (HR, 2.28 [95% CI, 1.46 to 3.57] vs HR, 0.74 [95% CI, 0.39 to 1.38]) compared with a reference group with AHI less than 20 and no excessive daytime sleepiness.

Three studies evaluated the association between AHI and cardiovascular mortality.\textsuperscript{12,150,151} Sample sizes ranged from 1522\textsuperscript{150} to 6294.\textsuperscript{151} Mean duration of follow-up ranged from 8.2\textsuperscript{151} to 13.8 years.\textsuperscript{150} Mean age ranged from 48\textsuperscript{150} to 63\textsuperscript{151} years. In multivariable analyses, all 3 studies reported that participants with severe or moderate to severe OSA at baseline had a higher risk of cardiovascular death (eFigure 29 in the Supplement), with HRs of 1.7 (95% CI, 1.1 to 2.5) (for men only in the Sleep Heart Health Study),\textsuperscript{151} 2.9 (95% CI, 1.1 to 7.3), and 5.9 (95% CI, 2.6 to 13.3).\textsuperscript{150} Variables included in the models are detailed in eTable 33 in the Supplement. Briefly, all of them included age, BMI, smoking, and multiple medical conditions or used matching for age and BMI. Two of 3 included alcohol use, blood pressure, and cholesterol level.

A single included study evaluated the association between AHI and the incidence of each of the following outcomes: cancer-
related mortality,145 nonfatal cardiovascular events,12 heart failure,146 coronary heart disease,148 stroke,145 cognitive impairment or dementia,146 and cognitive decline143 (eFigure 30 in the Supplement). Overall, findings for these outcomes were imprecise, consistency was unknown (with a single study for each), and evidence was often limited by risk of bias (especially risk of residual confounding).

Harms of Screening or Diagnostic Testing

Key Question 7a. Are there harms associated with screening or diagnostic testing for OSA?

Key Question 7b. Do the harms of screening or diagnostic testing differ for subgroups defined by age, sex, or BMI?

No eligible studies were identified.

Harms of Treatment

Key Question 8a. Are there harms associated with treatment of OSA?

Key Question 8b. Do the harms of treatment differ for subgroups defined by age, sex, BMI, or OSA severity?

Reporting of harms in the included studies was sparse. Twenty-two of the RCTs included in KQ4 reported harms associated with treatments for OSA: 9 trials of CPAP,66,70,75,88,91,101,105,108 8 of MADs,105,134-122 1 of a very low energy diet,132 4 of airway surgical treatments,123,124,126,128 and 1 of bariatric surgery (eTables 35 through 38 in the Supplement).125

Continuous Positive Airway Pressure

Of the 9 included RCTs, most enrolled fewer than 100 participants; 1 trial134 enrolled 281, and the Apnea Positive Pressure Long-term Efficacy Study (APPLES)70 enrolled 1098. Most of the studies followed up patients for 8 to 12 weeks. Overall, 2% to 47% of participants in trials reporting any harms had specific adverse events while using CPAP. In general, harms were likely short-lived and could be alleviated with discontinuation of CPAP or additional interventions. These harms included oral or nasal dryness, eye or skin irritation, rash, epistaxis, and pain.

Mandibular Advancement Devices

Of the 8 included RCTs,105,114-116,118,120-122 study durations ranged from 4 to 24 weeks. Across studies that reported any discontinuation because of adverse events, 7% of patients using MADs discontinued use, compared with 1% of control patients.105,116,122 The most commonly reported symptoms that occurred more often in active MAD study groups were oral dryness,105,114,115,122 excess salivation105,114,115,117,122 (although 1 study reported a higher rate of excessive salivation in the sham MAD group than in the active MAD group115), and oral mucosal, dental, or jaw symptoms.

Airway Surgery and Bariatric Surgery

Five trials reported harms of surgical treatment: 1 each of single-session soft palate radiofrequency surgery,123 temperature-controlled radiofrequency tissue ablation,128 uvulopalatopharyngoplasty,124 laser-assisted uvulopalatoplasty126 and bariatric surgery.125 Reported harms included postoperative bleeding; rehospitalization; difficulty speaking, breathing, drinking, opening the mouth, and swallowing; change in vocal quality; hematomas; ulcerations; infections; temporary nasal regurgitation; pain; and rehospitalization after bariatric surgery because of an acute proximal gastric pouch dilation that required additional surgery (eTable 38 in the Supplement).
Weight Loss, Diet, and Exercise Interventions
The single weight loss study that reported harms compared a very low energy diet with usual diet over 9 weeks. In the very low energy diet group, fewer than 10% of patients reported each of the following: constipation, dizziness, gout, and dry lips.

Discussion
The summary of findings is presented in Table 6. No eligible studies directly evaluated the effectiveness or adverse outcomes of...
Table 6. Summary of Evidence for Screening and Treatment of Obstructive Sleep Apnea (continued)

<table>
<thead>
<tr>
<th>Key Question and Topic</th>
<th>No. of Studies (Study Design)</th>
<th>No. of Participants</th>
<th>Summary of Main Findings (Including Consistency and Precision)</th>
<th>Applicability</th>
<th>Limitations (Including Reporting Bias)</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>5: Benefits of treatment for health outcomes</td>
<td>50 (RCTs)</td>
<td>6191</td>
<td>Evidence on most health outcomes was limited; too few trials reported them or too few events occurred to make conclusions for most health outcomes (eg, mortality, cardiovascular events, motor vehicle crashes). However, CPAP was associated with improvement in sleep-related quality of life, albeit with a small effect size (Cohen d, 0.28 [95% CI, 0.14 to 0.42]; 13 trials, 2325 participants).</td>
<td>Referral population with known OSA</td>
<td>Study durations may be insufficient to determine benefit for many health outcomes; small number of total events observed across studies (eg, for mortality, motor vehicle crashes, stroke, and cardiovascular events). Reporting bias not detected.</td>
<td>Fair</td>
</tr>
<tr>
<td>6: Association between AHI and health outcomes</td>
<td>11 (prospective cohort)</td>
<td>26 954</td>
<td>Increased risk of mortality for people with severe OSA (pooled HR, 2.07 [95% CI, 1.48 to 2.91]; 5 studies, 11 003 participants). Risk of cardiovascular mortality also increased (HRs from 2.9 [95% CI, 1.1 to 7.5] to 5.9 [95% CI, 2.6 to 13.3]). Findings were consistent and precise for all-cause mortality; consistent but imprecise for cardiovascular mortality.</td>
<td>General population</td>
<td>Risk of residual confounding (eg, due to physical activity, diet). Single study for all other outcomes (eg, cancer-related mortality, stroke, CHD). Reporting bias not detected.</td>
<td>Fair</td>
</tr>
<tr>
<td>7: Harms of screening or diagnostic testing</td>
<td>0 (NA)</td>
<td>NA</td>
<td>No studies were identified that directly evaluated the harms.</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>8: Harms of treatment</td>
<td>22 (RCTs)</td>
<td>2496</td>
<td>Common adverse effects of CPAP included oral or nasal dryness, eye or skin irritation, rash, epistaxis, and pain; common adverse effects of MADs included oral dryness, excess salivation, mucosal erosions, or pain (mucosal, dental, or jaw). Findings were consistent but imprecise for CPAP, inconsistent and imprecise for MADs, and of unknown consistency for other treatments (single study for each).</td>
<td>Referral population with known OSA</td>
<td>High heterogeneity in reporting and findings for CPAP and MADs. Single studies and small sample sizes for surgical interventions and weight loss programs. Reporting bias not detected.</td>
<td>Fair</td>
</tr>
</tbody>
</table>

Abbreviations: AHI, apnea-hypopnea index; AUC, area under the curve; CHD, coronary heart disease; CPAP, continuous positive airway pressure; DBP, diastolic blood pressure; ESS, Epworth Sleepiness Scale; HR, hazard ratio; MAD, mandibular advancement device; MVAP, multivariable apnea prediction; NA, not available; OSA, obstructive sleep apnea; OSAS, obstructive sleep apnea syndrome; RCT, randomized clinical trial; SBP, systolic blood pressure; WMD, weighted mean difference.

* The total number of trials reporting each outcome for CPAP is more than the number that contributed to the data in this column because the CPAP vs control data were not entered. Rather, the focus was on the CPAP vs sham data. However, evidence from both comparator groupings was considered in the assessments.

Screening compared with no screening. Potential harms include overdiagnosis and overtreatment for asymptomatic people (with AHI ≥5) who would never have developed symptoms of or problems from OSA, costs, and additional testing (eg, future polysomnographies to follow up patients over time). Furthermore, no eligible studies were found that evaluated the effect of screening on psychological outcomes such as distress due to labeling or stigma.

Very few eligible studies evaluated the accuracy of questionnaires or prediction tools for distinguishing people in the general population who are more or less likely to have OSA. The only screening approach with at least 2 included studies suggesting possible accuracy was the MVAP score followed by in-home portable monitoring for detecting severe OSAS. Although this approach may have potential for screening, the evidence was limited by potential spectrum bias, with oversampling of high-risk participants and those with OSA and OSAS, which may substantially overestimate the accuracy that would be achieved in the general population. Spectrum bias occurs when heterogeneity of test performance exists...
across subgroups and studies preferentially sample (intentionally or unintentionally) from a subgroup of the target population. None of the included studies evaluating MVAP prospectively measured calibration, often assessed by plotting the predicted risk vs an observed event rate, and none assessed clinical utility for improving health outcomes.

This review included fewer studies evaluating questionnaires or clinical prediction tools than some previously published reviews and guidelines, primarily because of the requirement that studies enroll asymptomatic adults or persons with unrecognized symptoms of OSA; referral populations (eg, to sleep clinics) were not eligible. The focus of previous reviews and guidelines was generally on diagnostic testing (of adults with symptoms suggestive of disordered sleep) rather than on screening (of asymptomatic people or those with unrecognized symptoms). Nevertheless, those reviews and guidelines generally reported low overall quality or strength of evidence for questionnaires and prediction tools.

Related to accuracy of diagnostic tests, there was limited evidence evaluating type II portable monitors. For type III and IV monitors, existing literature revealed some inconsistency, with wide ranges of sensitivity and specificity, especially for single-channel type IV monitors for detecting moderate to severe OSA. Nevertheless, many studies reported moderate to high positive likelihood ratios (>5) and moderate to low negative likelihood ratios (<0.2), and previous reviews and guidelines concluded that moderate-quality evidence shows that type III and IV monitors are “generally accurate to diagnose OSA, but have a wide and variable bias in estimating the actual AHI.” Evidence for type IV portable monitors is limited by inconsistency and imprecision. In addition, unlike other types of portable monitors, type IV monitors are limited by their inability to differentiate obstructive and central events.

This review found consistent evidence from RCTs that CPAP effectively reduces AHI to normal (<5) or near-normal levels, reduces excessive sleepiness, and reduces blood pressure. However, the clinical significance of mean reductions of 2 points on the ESS and 2 to 3 points for blood pressure measures is somewhat uncertain. For sleepiness, the data suggest a clinically significant reduction in most included trials because 85% of the trials in the meta-analysis for ESS that had mean baseline ESS scores of 10 or greater (indicating excessive daytime sleepiness) reported mean end-point ESS scores in the normal range of less than 10162,163 for the CPAP groups (mean end-point ESS score across studies was <8). However, the threshold for a clinically significant change in ESS score is somewhat uncertain. Although a recent review noted that experts consider a 1-point change clinically significant,16 other sources suggest that a greater change, of at least 3 or 4 points, should be the clinically significant threshold. For example, some trials that use ESS score as an outcome have considered a change of 4 or more points to be clinically significant for their sample size calculations or interpretation of findings.162,164 Regardless of what constitutes a clinically significant change, potential bias from the subjective nature of the ESS remains, and some authors have raised concerns about its construct validity (ie, uncertainty regarding whether it is an accurate measure of sleepiness).165-167

For blood pressure reduction, some authors suggest that a difference of more than 9/10 (systolic/diastolic) mm Hg is clinically meaningful for individuals.168-170 However, across a population, guidelines have suggested that reductions of 2 to 3 mm Hg for systolic blood pressure could result in a significant reduction in cardiovascular mortality (by 4%-5% for coronary heart disease and 6%-8% for stroke).171

MADs and weight loss programs also reduce AHI and excessive sleepiness, although the magnitudes of effects were generally less than with CPAP, and blood pressure reduction was not established. Although this review did not evaluate head-to-head studies (eg, directly comparing MADs with CPAP), previous comparative effectiveness reviews examining head-to-head trials reported smaller effect sizes for MADs than for CPAP for reducing AHI.171 Evidence on surgical treatments was limited by unknown consistency and precision, because only a single RCT evaluated each surgical technique studied.

Evidence on most health outcomes was limited; too few RCTs reported them or too few events occurred to make conclusions about the effectiveness for reducing mortality, cardiovascular events, or motor vehicle crashes. However, the meta-analysis for sleep-related quality of life found a significant benefit for CPAP, albeit with a small effect size.

Reporting of harms from treatment in the included studies was sparse. In general, the adverse events related to CPAP treatment were likely short-lived and could be alleviated with discontinuation of CPAP or additional interventions. No included studies reported on psychosocial harms of treatment, such as marital stress due to disruption of partner sleeping (eg, because of the noise of CPAP).

Adverse effects may limit adherence to treatment. A wide range of adherence to CPAP usage recommendations has been reported, ranging from about 30% to 85%.172 A systematic review reported that cohort studies with multivariable analyses for predictors of nonadherence showed that 14% to 32% of patients discontinued CPAP over 4 years and that patients used CPAP for an average of 5 hours per night; data were too limited to provide adherence rates for MADs.21 A recent Cochrane systematic review of 33 studies (2047 participants) found low- to moderate-quality evidence that 3 types of interventions can increase CPAP usage in CPAP-naive participants.172 However, they noted that trials did not assess people who have struggled to adhere, and the effect of improved CPAP usage on health outcomes remains unclear.

Consistent evidence from prospective cohort studies that focused on community-based participants supports the association between AHI and all-cause mortality. People with severe (AHI ≥30) or moderate to severe (AHI ≥15) OSA had a hazard ratio for death of 2.07 compared with controls when pooling data from multivariable analyses. There was also consistent evidence showing that people with severe or moderate to severe OSA have increased cardiovascular mortality. The cohort studies controlled for many potential confounders, but residual confounding attributable to health-related factors associated with OSA (eg, physical activity, diet) and generally not accounted for is possible.

This review had limitations. The ability to describe the direct evidence on the effectiveness or harms of screening was inadequate, because no studies comparing screened and unscreened populations were identified. Therefore, literature was reviewed that might establish an indirect chain of evidence from multiple questions that link screening to health outcomes. For the first question in that indirect pathway (KQ2), there was limited evidence that one screening approach might be useful to screen for severe OSAS, but the evidence was limited by potential spectrum bias, and no studies
prospectively assessed calibration or clinical utility for improving health outcomes. In addition, this review did not evaluate the accuracy of individual physical examination findings. Questionnaires or clinical prediction tools were required to have multiple factors because previous systematic reviews have found limited utility of individual findings. A recent review of clinical examination accuracy, which was not limited to asymptomatic patients or those with unrecognized symptoms, found that (among individual symptoms or signs) the most useful observation for identifying patients with OSA was nocturnal choking or gasping, imparting a small increase in the likelihood of disease (summary likelihood ratio, 3.3 [95% CI, 2.1 to 4.6]) when the diagnosis was established by AHI ≥10.1 The review found that many symptoms and signs provide limited information in determining the likelihood of OSA.1

Conclusions

There is uncertainty about the accuracy or clinical utility of all potential screening tools. Multiple treatments for OSA reduce AHI, ESS scores, and blood pressure. Trials of CPAP and other treatments have not established whether treatment reduces mortality or improves most other health outcomes, except for modest improvement in sleep-related quality of life.

REFERENCES


obstructive sleep apnoea: randomised controlled trial. BMJ. 2010;341:c5991.
101. Engleman HM, Kingshott RN, Wraith PK, Mackay TW, Deary IJ, Douglas NJ. Randomised


