Systematic review of the literature regarding the diagnosis of sleep apnea
Ross S D, Allen I E, Harrison K J, Kvasz M, Connelly J, Sheihait I A

Authors' objectives
To investigate the evidence for diagnosing sleep apnoea (SA) in adults.

Searching
MEDLINE and Current Contents were searched from 1980 to November 1997. The search terms were reported. In addition, the bibliographies of all included studies and relevant review articles were checked for any additional articles. Studies published in English, German, French, Spanish or Italian were eligible.

Study selection
Study designs of evaluations included in the review
Reviews, meta-analyses and case-studies were excluded. Only studies of at least 10 patients were included. The included studies were mainly case series and observational studies.

Specific interventions included in the review
Studies of any test to establish or support a diagnosis of SA were included. Studies of treatments were excluded. The included studies used the following tests: portable sleep monitoring devices, partial channel polysomnography (PSG), oximetry, partial time PSG, radiologic imaging (magnetic resonance imaging, cephalometry, computed tomography) and clinical tests (flow volume loops, global impressions, prediction equations, laboratory assays, and focused questionnaires).

Reference standard test against which the new test was compared
Only studies that compared a test relative to a standard sleep laboratory PSG-derived apnoea index (AI: the number of apnoeic episodes/hour sleep), apnoea-hypopnoea index (AHI: total apnoeas plus hypopnoeas during total time sleep, divided by the number of hours asleep) or respiratory distress index (RDI) were included. The included studies varied with regard to the criteria used to constitute a standard PSG. All of the studies included measures of sleep: electroencephalogram, electrooculogram or submental electromyogram (EMG). Other measures used were respiratory activity, oxygen saturation (oximetry), cardiac arrest (electrocardiogram), tibial EMG, snoring and/or body position. In the included studies, standard PSGs were performed in a sleep laboratory. The thresholds for AI or AHI (or RDI) ranged from 5 to 40.

Participants included in the review
Studies of adults suspected of or diagnosed with any form of SA (obstructive, central, mixed or unspecified) were included. Studies of special populations (e.g. neuromuscular diseases or cerebral malformations, congenital or acquired structural abnormalities of the head or neck) were excluded. The mean age of the participants in the included studies was 49 years (range: 36 to 60).

Outcomes assessed in the review
Studies that reported the outcomes in terms of sensitivity and specificity (or a function of these outcomes, i.e. likelihood ratios), or that reported sufficient information to calculate these or a correlation between the new test and the diagnosis of obstructive SA by full PSG, were included. Studies in which the outcomes were not extractable (i.e. where results for patients with other potentially confounding diseases could not be separated from the results for SA patients) were excluded. The lowest AI or AHI threshold for SA diagnosis was used to determine the sensitivity and specificity.

How were decisions on the relevance of primary studies made?
The authors did not state how the papers were selected for the review, or how many reviewers performed the selection.

Assessment of study quality
The quality of the studies was rated for diagnostic test study design, execution and reporting, using a rating instrument. The quality criteria were reported in full. The scores could range from 0 to 44. The authors did not state how the papers were assessed for quality, or how many reviewers performed the quality assessment.

**Data extraction**
Two reviewers independently extracted the data from the studies. Any differences were resolved by consensus, or by a third reviewer if necessary. One reviewer was blinded to the source of funding, authors and journal. Data relating to study level, patient level, test characteristics and clearly reported aggregate results were extracted. Studies with a validity score of less than 16 were not data extracted.

**Methods of synthesis**

**How were the studies combined?**
The Mantel-Haenszel fixed-effect model was used to combine the comparative summary statistics for groups according to diagnostic test category. In addition, study and patient-level covariates and study evidence scores were also summarised. Where data were available, a summary receiver operating characteristic (ROC) curve was calculated for each diagnostic group. Studies with a validity score of less than 16 were excluded from the analysis.

**How were differences between studies investigated?**
Heterogeneity was assessed using the summary ROC curves. The impact of the covariates on the summary ROC model was investigated through a sensitivity analysis.

**Results of the review**
A total of 71 diagnostic or screening studies (n=7572) were included.

All studies: the mean evidence score was 20.6 (range: 16 to 34).

Partial channel PSGs (3 studies, n=213): the mean evidence score was 17.7 (range: 17 to 19). The sensitivity ranged from 82 to 94% and the specificity from 82 to 100%.

Portable devices (25 studies, n=1,631): the mean evidence score was 22.1 (range: 16 to 34). The results were mainly from supervised sleep laboratories. Reliability in unattended home use, equipment failure rates, night-to-night reproducibility, compliance, safety and price were seldom reported. The sensitivity ranged from 32 to 100% and the specificity from 33 to 100%. Study and device heterogeneity were apparent.

Oximetry (12 studies, n=1,784): the mean evidence score was 20 (range: 16 to 32). The sensitivity ranged from 36 to 100% (pooled estimate, PE=87.4%, standard error, SE=3.8), and the specificity from 23 to 99% (PE=64.9%, SE=6.7). The ROC curve indicated little heterogeneity.

Partial time PSGs (7 studies, n=505): the mean evidence score was 18.6 (range: 17 to 20). The sensitivity ranged from 66 to 93% (PE=69.7%, SE=5.3) at an AI or AHI threshold of 5, and from 42 to 89% (PE=79.5%, SE=5.2) at a threshold of 10. The specificity ranged from 50 to 100% (PE=87.4%, SE=5.4) at an AI/AHI threshold of 5, and from 57 to 100% (PE=86.7%, SE=4.6) at a threshold of 10. The ROC curve indicated that, with the exception of one study that had low sensitivity and extremely high specificity, all the studies were homogeneous.

Flow volume loops (4 studies, n=595): the mean evidence score was 18.3 (range: 17 to 20). The sensitivity and specificity ranged from 41 to 59% (mean 39.1%) and from 54 to 85% (mean 60.5%), respectively, when both a measure of extrathoracic airway obstruction and a measure indicative of pharyngeal fluttering during respiration were analysed together.

Clinician global impressions (4 studies, n=1,139): the mean evidence score was 23 (range: 19 to 28). The sensitivity ranged from 52 to 79% (mean 58.9%), and the specificity from 50 to 100% (mean 65.6%). The ROC curve indicated that the sensitivity was relatively constant across studies, but there was much variation in the specificity.
There were 5 radiology studies and 9 clinical studies (concerning anthropomorphic signs or ears, nose, throat exams) that could not be analysed due to insufficient data.

Prediction equations (8 models, n=1,908): the mean evidence score was 21.5 (range: 17 to 30). The sensitivity ranged from 28 to 98% (mean 66.5%), and the specificity from 21 to 100% (mean 88.7%).

**Authors' conclusions**
The authors concluded that the available evidence suggests that full PSG is best for diagnosing SA. There has been some progress in establishing the sensitivity and specificity of other tests, and future research should build on this evidence base. The standardisation of diagnostic criteria and terms is necessary to further development and improve the usefulness of future literature.

**CRD commentary**
The objective of the review was clearly stated and the predetermined inclusion and exclusion criteria were appropriate. By limiting the search to published literature, relevant studies may have been missed and publication bias introduced. Relevant literature may also have been omitted by the exclusion of studies in languages other than those considered (English, German, French, Spanish and Italian). The quality assessment, which was derived from a published checklist, was appropriate. The studies were appropriately synthesised and heterogeneity were assessed. Details of the included studies were tabulated clearly, but it would have been useful if the authors had provided details of the study design. Two reviewers carried out the data extraction independently, thus minimising potential bias in this process. However, since there were no details of how the study selection and quality assessment were conducted, it was unclear if steps were taken to minimise bias in these processes. The authors' conclusions follow on from the findings.

**Implications of the review for practice and research**
**Practice:** The authors did not state any implications for practice.

**Research:** The authors stated that future studies should use common definitions and terminology for terms such as apnoea and hypopnoea. Research should establish the relations between AHI and AI, and clarify what the frequency of sleep apnoea/hypopnoea is in general populations, by age and gender. Further research on sleep studies in the home should be carried out, and long-term studies reporting the results of treated SA versus untreated SA are required. Research is also required to validate all sleep monitoring systems, proposed as replacements or prequalifiers for PSG, in settings in which they are intended to be used.

**Funding**
Agency for Health Care Policy and Research, contract number 290-97- 0016.

**Bibliographic details**

**Original Paper URL**
http://www.ahrq.gov/clinic/epcsums/apneasum.htm

**Other publications of related interest**
This additional published commentary may also be of interest. Douglas N. Review: screening tests are not as accurate as overnight polysomnography for the diagnosis of adult sleep apnoea. Evid Based Med 2000;5:61.

**Indexing Status**
Subject indexing assigned by CRD

MeSH
Sleep Apnea, Obstructive /diagnosis

AccessionNumber
12000008301

Date bibliographic record published
31/03/2004

Date abstract record published
31/03/2004

Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.