Proposed Decision Memo for Continuous Positive Airway Pressure (CPAP) Therapy for Obstructive Sleep Apnea (OSA) (CAG-00093R)

Decision Summary

The Centers for Medicare & Medicaid Services (CMS) is seeking public comment on the following proposed conclusions:

The evidence is not adequate to conclude that the use of unattended portable multi-channel sleep testing with a minimum of 7 monitored channels including EEG, EOG, EMG, ECG or heart rate, airflow, respiratory effort, and oxygen saturation (Type II Devices based on the 1994 ASDA classification) is reasonable and necessary in the diagnosis of OSA and these tests will remain noncovered for this purpose.

The evidence is not adequate to conclude that the use of unattended portable multi-channel sleep testing with a minimum of 4 monitored channels including ventilation or airflow, heart rate or ECG, and oxygen saturation (Type III Devices based on the 1994 ASDA classification system) is reasonable and necessary in the diagnosis of OSA and these tests will remain noncovered for this purpose

CMS is requesting public comments on this proposed decision memorandum pursuant to Section 731 of the Medicare Modernization Act. After considering the public comments and any additional evidence, we will make a final determination and issue a final coverage decision memorandum

Back to Top

Proposed Decision Memo

To: Administrative File: CAG 00093R, Continuous Positive Airway Pressure (CPAP) Therapy for Obstructive Sleep Apnea (OSA)

From:

Steve E. Phurrough, MD, MPA
Director, Coverage and Analysis Group

Louis B. Jacques, MD
Director, Division of Items and Devices

LCDR Tiffany Sanders, MD
Lead Medical Officer
Division of Items and Devices

Francina Spencer
Lead Analyst
Division of Items and Devices

James Rollins, MD, PhD
Medical Officer
Division of Items and Devices

Jackie Sheridan-Moore
Analyst
Division of Items and Devices

Subject: Proposed Decision: CPAP Therapy for OSA

Date: January 7, 2005

I. Proposed Decision

The Centers for Medicare & Medicaid Services (CMS) is seeking public comment on the following proposed conclusions:

The evidence is not adequate to conclude that the use of unattended portable multi-channel sleep testing with a minimum of 7 monitored channels including EEG, EOG, EMG, ECG or heart rate, airflow, respiratory effort, and oxygen saturation (Type II Devices based on the 1994 ASDA classification) is reasonable and necessary in the diagnosis of OSA and these tests will remain noncovered for this purpose.

The evidence is not adequate to conclude that the use of unattended portable multi-channel sleep testing with a minimum of 4 monitored channels including ventilation or airflow, heart rate or ECG, and oxygen saturation (Type III Devices based on the 1994 ASDA classification system) is reasonable and necessary in the diagnosis of OSA and these tests will remain noncovered for this purpose

CMS is requesting public comments on this proposed decision memorandum pursuant to Section 731 of the Medicare Modernization Act. After considering the public comments and any additional evidence, we will make a final determination and issue a final coverage decision memorandum

II. Background

On April 8, 2004, CMS began a national coverage determination process for the diagnosis of patients with obstructive sleep apnea (OSA) requiring continuous positive airway pressure (CPAP) therapy. Current national coverage guidelines specify that only polysomnography done in a facility-based sleep study laboratory be used to identify patients with OSA requiring CPAP (National Coverage Decision Manual Section 240.4) (Formerly CIM 60-17). CMS has received a request from Dr. Terence M. Davidson, MD, of the University of California San Diego, School of Medicine to modify this decision to include the use of portable multi-channel home sleep testing devices as an alternative to facility-based polysomnography in the evaluation of OSA.

Sleep apnea refers to a collection of conditions and syndromes that are characterized by periods of apnea, a temporary cessation of breathing. It was initially described in the early 1800's. One of the first accounts was written by Charles Dickens in 1837 and entitled *The Posthumous Papers of the Pickwick Club*. Subsequently, William Osler in 1918 coined the term "Pickwickian" to describe the obese, hypersomnolent patient. Over the years, various sleep apnea syndromes have been described and classified into three main types: central, obstructive, and mixed. Central sleep apnea refers to apnea syndromes with origins in the central nervous system. Obstructive sleep apnea (OSA) refers to apnea syndromes due primarily to collapse of the upper airway during sleep. Mixed apnea refers to apnea with both central and obstructive characteristics. Of the three main types of apneas, OSA has received the most scientific interest and study. The prevalence of OSA in the United States has been estimated to be about 2-4% of middle aged adults.[1]

OSA has also been identified as a risk factor for other medical conditions including hypertension, nocturnal cardiac arrhythmias, cerebrovascular accidents, and myocardial infarctions.[2] The pathogenesis and pathophysiology of OSA have been studied extensively. During sleep, the upper airway becomes occluded, resulting in an episode of apnea. As a result of the apnea, the patient experiences a brief arousal from sleep. With the return of breathing, the patient typically returns to sleep quickly. This sequence occurs repeatedly. The pharynx has been identified as the primary site of obstruction in most patients. A number of anatomical and functional factors, such as negative oropharyngeal pressure, decreased muscle activity, and possible narrowing of the oropharyngeal lumen may also be involved in the collapse of the upper airway during sleep.

Symptoms of OSA include somnolence, fatigue, irritability, headaches, cognitive impairment, depression, and personality changes.[3] There are a number of medical and surgical treatment options for OSA.[4] Nonpharmacologic medical treatments include education regarding sleep hygiene, weight reduction, tongue-retaining devices, positive airway pressure modalities such as continuous positive airway pressure, and bi-level positive airway pressure (BiPAP). CPAP involves the administration of air usually through the nose by an external device at a fixed pressure to maintain the patency of the upper airway. Medications that may be used in OSA include oxygen, protriptyline and theophylline. Surgical procedures include uvulopalatopharyngoplasty, somnoplasty and tracheostomy.

Laboratory based polysomnography, with continuous overnight monitoring of various neurophysiologic and cardiorespiratory parameters of sleep, has been the mainstay in the diagnostic work-up for persons suspected of having OSA. Polysomnography monitors sleep stages, respiratory effort, oxygen saturation, heart rate, body position, and limb movements. From the collected data, measurements such as the apnea hypopnea index (AHI) can be calculated and used to diagnose specific sleep disorders.

While the occurrence of apnea has remained a primary diagnostic criterion for sleep apnea, episodes of reduced ventilation have received considerable attention and clinical consideration since the 1980's. The term hypopnea has been used to describe these episodes of reduced breathing; however, there was no general consensus for the definition of hypopnea at the time.[5] Variations in the definition of hypopnea still persist today. Despite such variations, the apnea-hypopnea index (AHI = number of episodes of apneas and hypopneas per hour of sleep) has been utilized extensively in recent years in the published literature in the definition of OSA. The AHI has also been called the respiratory distress index (RDI).

Over the past several years, a number of portable devices have been developed that measure to varying extents similar neurophysiologic and cardiorespiratory parameters of sleep as those obtained with laboratory based polysomnography. In 1994, the American Sleep Disorders Association developed a classification system for these devices.

Type I devices are considered standard laboratory-based polysomnography. Type II devices are comprehensive portable polysomnographic devices with a minimum of seven channels which measure the same neurophysiologic and cardiorespiratory parameters of sleep as standard polysomnography. These devices allow for the measurement of sleep staging. Type III devices have a minimum of four channels and measure only cardiorespiratory parameters of sleep. Because these devices do not permit the determination of sleep versus wakefulness, abnormal breathing events are calculated as "events per hour in bed" instead of "events per hour of sleep." Type IV devices measure only one or two respiratory parameters such as oxygen saturation or airflow.

III. History of Medicare Coverage

In 1986, CMS (then known as the Health Care Financing Administration) requested the Office of Health Technology Assessment (OHTA) to conduct an assessment of the safety, clinical effectiveness and use of CPAP. OHTA reported that "the consensus of clinical opinion from the available information appears to be that CPAP can in the majority of cases prevent OSA and provide substantial clinical improvement with minimal associated morbidity." They went on further to recommend that "the use of CPAP be covered under Medicare when used in adult patients with moderate and severe OSA who have failed to obtain relief from other non-invasive therapies and for whom surgery would be the only other therapeutic alternative."[6] The diagnosis of OSA required at least 30 episodes of apnea, each lasting a minimum of 10 seconds, during 6-7 hours of sleep. These specifications were based predominately on expert opinions at the time.[7]

Based on the OHTA technology assessment, Medicare issued an NCD (see NCD Manual 240.4) which covered CPAP for adult patients with moderate or severe OSA for whom surgery is a likely alternative (effective date January 12, 1987), and adopted OHTA's recommendations on the diagnosis of OSA. Since the 1986 decision specifically addressed CPAP only, the Durable Medical Equipment Regional Carriers (DMERCs) have issued a respiratory assist devices regional medical review policy (RAD RMRP) that addresses BiPAP devices and other accessories (last revised in 1999). Specifically for the treatment of OSA, a respiratory assist device with bilevel pressure capability, without backup rate feature, used with noninvasive interface will be covered for the first three months of noninvasive positive pressure respiratory assistance (NPPRA) if the following criteria are met:

- complete facility-based, attended polysomnogram has established the diagnosis of obstructive sleep apnea, and
- single level device (CPAP) has been tried and proven ineffective.

Unattended home sleep study testing has been under review by CMS since 1989. The latest review occurred in 1995. In 1995, the agency's reviewing body for the development of national coverage determinations (formerly the Technical Advisory Committee) concluded that the safety and effectiveness of home studies used to diagnosis sleep disorders was unproven and thus should not be covered by the Medicare program. The TAC recommended that this issue be reconsidered for national policy following the completion of a large study of sleep disorders by the NIH. This was to include an evaluation of in-home testing. The study was expected to be completed within two to three years. Therefore, the coverage of unattended home sleep study testing was left to carrier discretion.

In 2001 the national coverage policy on CPAP was expanded to include Medicare beneficiaries with an apnea/hypopnea index (AHI) of $\geq$ 15, or an AHI $\geq$ 5 and $\leq$ 14 with documented symptoms of excessive daytime sleepiness, impaired cognition, mood disorders or insomnia, or documented hypertension, ischemic heart disease or history of stroke. However, the guidelines specified that only a polysomnography done in a facility-based sleep study laboratory could be used to identify patients with obstructive sleep apnea.

In 2003 Medicare spent $177 million for CPAP therapy and related supplies and services. In addition, Medicare spent $35 million for unattended sleep study testing.

IV. Timeline of Recent Activities

April 8, 2004	Request posted and the beginning of the initial 30-day comment period on this NCD for scientific input relevant to the issue under consideration.
April 13, 2004	A Benefit Category Determination (BCD) was requested from the Center for Medicare Management (CMM).
May 27, 2004	The BCD was approved by CMM.
June 25, 2004	
	Comments from the initial comments period were posted. The public was invited to participate in a second 30-day period. Comments were requested on the following questions: How does the diagnostic test performance of unattended portable multi-channel home sleep testing compare to facility-based polysomnography in the diagnosis of obstructive sleep apnea? a. If unattended portable multi-channel home sleep testing is as effective as polysomnography in the diagnosis of obstructive sleep apnea which parameters of sleep and cardiorespiratory function (i.e. sleep staging, body position, limb movements, respiratory effort, airflow, oxygen saturation, ECG) are required? b. If unattended portable multi-channel home sleep testing is as effective as polysomnography in the diagnosis of obstructive sleep apnea what conditions (i.e. patient education, technician support) are required so that it is done correctly in the home?
July 2, 2004	Requested a Technology Assessment from the Agency for Healthcare Research and Quality.
July 29, 2004	Announced the presentation of this issue to the Medicare Coverage Advisory Committee (MCAC).
August 27, 2004	Federal Register Notice published announcing MCAC. Instructions for presenters are given in the Federal Register Notice.
August 31 -June 3, 2004	CMS held multiple meetings with industry representatives. Information from industry representatives, related articles from Medline searches, and public comments were obtained and reviewed.
September 2, 2004	The MCAC panel questions posted for review.
September 7, 2004	The Technology Assessment Report, second round of comments, and the MCAC Roster posted.

September 28, 2004	The issue was presented to the MCAC.

V. FDA Status

These and other similar devices, such as multi-channel home sleep study testing and other related devices have been considered and cleared for marketing by the Food and Drug Administration (FDA) under a 510(k) process. The 510(k) is a notification of intent to market a specific device. The FDA has determined that certain home sleep study testing devices are "substantially equivalent to legally marketed predicate devices marketed in interstate commerce prior to May 28, 1976, enactment date of the Medical Device Amendments, or to devices that have been reclassified in accordance with the provisions of the Federal Food, Drug, and Cosmetic Act." A substantially equivalent determination assumes compliance with the Good Manufacturing Practice requirements, as set forth in the Quality System Regulation (QS) for Medical Devices: General regulation (21 CFR Part 820) and that, through periodic QS inspections, the FDA will verify such assumptions. Failure to comply with the GMP regulation may result in regulatory action. Typically, no clinical data is required as part of the 510 (k) application, but instead the clearance process focuses on technical performance. However, the FDA does request clinical data for snore validation as well as event detection (i.e. clinical validation that the apneas or hypopneas detected are also scored as apneas or hypopneas by a manual scorer). The FDA also compares sensitivity and positive predictive values to a predicate device.

The FDA has cleared many devices that allow a patient to wear a device that collects airflow and other patient measurements into a device that records data. The patient then takes the device to the physician and the physician downloads information that determines whether the patient has apnea sleep-related breathing disorder including obstructive sleep apnea or needs further sleep studies or assessment. There are currently many sleep assessment devices on the market cleared by the FDA through the 510(k) process for use in the home.

VI. General Methodological Principles

When making national coverage determinations, CMS evaluates relevant clinical evidence to determine whether or not the evidence is of sufficient quality to support a finding that an item or service is reasonable and necessary for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member (§1862(a)(1)(A) of the Social Security Act.) The critical appraisal of the evidence enables us to determine to what degree we are confident that: 1) the specific assessment questions can be answered conclusively; and 2) the intervention will improve net health outcomes for patients. The general methodological principles of study design utilized in our review of the evidence are in Appendix B.

VII. Evidence

We are providing a summary of the evidence we have considered to date. We will, of course, consider additional evidence submitted through the public comment period.

A. Introduction

Consistent findings across studies of net health outcomes associated with an intervention or diagnostic test as well as the magnitude of its risks and benefits are key considerations in the coverage determination process. For this decision memorandum, CMS held a Medicare Coverage Advisory Committee (MCAC) meeting and commissioned an external technology assessment (TA) from the Agency for Healthcare Research and Quality (AHRQ) to review published clinical evidence on the use of unattended portable monitoring devices in the diagnosis of OSA. CMS reviewed information and recommendations provided as a result of the MCAC meeting, the TA provided by AHRQ, and an independent search and review of individual clinical studies addressing this issue. We also received information from professional societies and other groups/organizations, searched evidence based practice guidelines, consensus statements, and position papers.

Outcomes of interest for a diagnostic test are not limited to determining its accuracy but include beneficial or adverse clinical effects, such as change in management due to test findings or preferably, improved health outcomes for Medicare beneficiaries. Accuracy refers to the ability of the test to distinguish patients who have or do not have the target disorder when compared to a reference standard. Measures used to determine accuracy include sensitivity (probability of a positive test result in patients with disease) and specificity (probability of a negative test in patients who do not have the disease).

In evaluating diagnostic tests based on a reference standard, comparable sensitivity and specificity values would be an outcome of interest. In the absence of direct evidence to show that the diagnostic test under review improves health outcomes, evidence of improved sensitivity or specificity could still prove useful as an intermediate outcome and data point estimate in the construction or a decision.

There is no anatomic or physiologic “gold standard” for the diagnosis of obstructive sleep apnea, in contrast to conditions such as cancer where a tissue biopsy result is the definitive standard reference. In studies that compare portable home sleep monitoring to facility-based polysomnography (PSG) performed in a sleep laboratory, the investigators have used the PSG result as the standard reference, i.e. the PSG result is used to define the true disease state for the individual patient. This is less than ideal, but represents the practical difficulty in diagnosing obstructive sleep apnea (OSA). Given the absence of a true “gold standard” reference, the clinical application of terms such as sensitivity and specificity is not straightforward.

Such evidence permits only the comparison of home sleep monitoring to facility-based PSG. It is problematic to make the inferential leap from there to a judgment on the ability of home sleep monitoring or PSG to accurately identify those patients who will, if untreated with CPAP, suffer the morbidity and mortality of obstructive sleep apnea. If an individual patient has conflicting results with these two tests, e.g. a negative home test in the face of a positive PSG, there is no available higher reference to determine whether the conflict arises from a false negative home test or a false positive PSG.

B. Discussion of evidence reviewed

1. Assessment questions

The development of an assessment in support of Medicare coverage decisions is based on the same general question for almost all requests: "Is the evidence sufficient to conclude that the application of the technology under study will improve net health outcomes for Medicare patients?"

The formulation of specific questions for the assessment recognizes that the effect of an intervention can depend substantially on how it is delivered, to whom it is applied, the alternatives with which it is being compared, and the delivery setting. In order to evaluate the net health outcomes of using unattended portable multi-channel sleep monitoring devices for the diagnosis of OSA as compared to laboratory based polysomnography, CMS sought to address the following questions:

Question 1: How does the diagnostic test performance of unattended portable multi-channel home sleep testing devices compare to facility-based polysomnography in the diagnosis of OSA?

a. If unattended portable multi-channel home sleep testing is as effective as polysomnography in the diagnosis of obstructive sleep apnea, which parameters of sleep and cardiorespiratory function (i.e. sleep staging, body position, limb movements, respiratory effort, airflow, oxygen saturation, electrocardiogram) are required?

b. If unattended portable multi-channel home sleep testing is as effective as polysomnography in the diagnosis of obstructive sleep apnea, what conditions (i.e. patient education, technician support) are required so that it is done correctly in the home?

2. External technology assessments

Systematic reviews are based on a comprehensive search of published studies to answer a clearly defined and specific set of clinical questions. A well-defined strategy or protocol (established before the results of the individual studies are known) guides this literature search. Thus, the process of identifying studies for potential inclusion and the sources for finding such articles is explicitly documented at the start of the review. Finally, systematic reviews provide a detailed assessment of the studies included.[8]

CMS commissioned a technology assessment from AHRQ to assess the utility of unattended portable monitoring devices in the diagnosis of OSA.[9, 10] This TA was an update to a systematic review originally published in 2003.[11] The following is a summary of the TA search strategy and findings.

Search strategy
A search of the MEDLINE database, The Cochrane Library, the National Guidelines Clearinghouse, and the International Network of Agencies for Health Technologies Assessment (INAHTA) database and a hand search of bibliographies included in the articles were conducted. This TA specifically searched for and evaluated literature published since 2002. Filters and limitations were used, and inclusion and exclusion criteria were developed to identify articles to be reviewed. One hundred seventy-two unique titles and abstracts were identified. One hundred fifty-seven articles did not meet inclusion criteria. Fifteen articles were retrieved for full review and 12 met inclusion criteria and were reviewed in detail. Four studies evaluated Type III devices. Only two[12] of these four compared results of portable device studies performed in the home with laboratory based polysomnography.

Results and appraisal
Three of the four studies evaluating Type III devices were rated as either fair or poor in terms of the quality of the evidence. The percentage of patients with missing data was in the range of 13-18% for unattended home studies. Sensitivities for unattended home studies were 91-95% and specificities were 81-91% for AHI > 15. Sensitivity and specificity values for other AHI cut-off points were noted to be similar. Studies reporting agreement measures such as correlation coefficients or Bland-Altman plots, noted good agreement for results obtained with the portable monitoring devices as compared to polysomnography. Manual scoring of portable monitoring device results was noted to be more discriminate in calculating AHI as compared to automated scoring.

The TA authors conclude that most articles only provided information on the use of portable monitoring devices in the laboratory setting when performed simultaneously with polysomnography. These studies do not provide information on the use and performance of portable monitoring devices unattended in the patients' home. The studies evaluated reported a wide range of data loss results for in home studies. The literature reviewed suggests that data loss appears to be greater when the patient performs set-up of the equipment. Results obtained from automated portable device scoring appear to provide less agreement with polysomnography than does manual scoring. "More evidence is needed to reach conclusions about the effect of co-morbidities, age, patient versus technician performed hookup on the overall effectiveness of home studies in diagnosing OSA compared to in-laboratory PSG."

3. Internal technology assessments

Search Strategy
An initial search of the MEDLINE® database was conducted on April 18, 2004. This search was updated on December 2, 2004. Filters and limitations were used, and inclusion and exclusion criteria developed to identify articles to be reviewed. The search used applicable MeSH heading and text words. Articles providing information regarding technical feasibility only were excluded from further review. Articles pertaining to devices that included only 1 or 2 channels of physiologic information to define sleep disordered breathing events (Type IV devices based on the 1994 American Sleeps Disorders Association classification) were also excluded from further review. In addition, the requestor provided abstracts or citations for 26 articles. Articles pertaining to Type IV devices, evaluating the use of auto-titration of CPAP, poster presentations, and those not in the English language were excluded from review. All articles identified and reviewed during our internal technology assessment, were also reviewed as part of the initial or updated external TA.

This section summarizes the findings of the systematic review performed by CMS on the use of unattended portable multi-channel home sleep testing devices in the diagnosis of OSA. It includes a summary of the results of 21 articles. For discussion purposes, studies are grouped by device type: (1) those with a minimum of 7 monitored channels including EEG, EOG, EMG, ECG or heart rate, airflow, respiratory effort, and oxygen saturation (Type II Devices based on the 1994 ASDA classification system) and (2) those with a minimum of 4 monitored channels including, ventilation or airflow, heart rate or ECG, and oxygen saturation (Type III Devices based on the 1994 ASDA classification system). Devices that included only 1 or 2 channels of physiologic information to define sleep disordered breathing events (Type IV Devices based on the ASDA classification system) were not considered multi-channel devices and were not reviewed as part of this decision. For a detailed description of each article, refer to the evidence tables provided under Appendix A.

Question 1: How does the diagnostic test performance of unattended portable multi-channel home sleep testing devices compare to facility-based polysomnography in the diagnosis of OSA?

a. If unattended portable multi-channel home sleep testing is as effective as polysomnography in the diagnosis of obstructive sleep apnea, which parameters of sleep and cardiorespiratory function (i.e. sleep staging, body position, limb movements, respiratory effort, airflow, oxygen saturation, electrocardiogram) are required?

b. If unattended portable multi-channel home sleep testing is as effective as polysomnography in the diagnosis of obstructive sleep apnea, what conditions (i.e. patient education, technician support) are required so that it is done correctly in the home?

Five studies[13] were reviewed that addressed the diagnostic test performance of portable multi-channel home sleep testing devices with a minimum of seven monitored channels including EEG, EOG, EMG, ECG or heart rate, airflow, respiratory effort, and oxygen saturation. Laboratory based polysomnography was used as the reference standard. The number of participants ranged from 20 to 103. Study participants were predominantly male with a mean age in the range of 45-52.

Portier (2000) studied 103 patients referred to a sleep laboratory for work-up of possible OSA. Patients underwent both an unattended portable study in the home and a laboratory based study. Minisomno, the portable device utilized, was described as being able to collect and store 8 hours of data from 10 to 18 channels. Patients came into the sleep laboratory for set-up of the portable device. A total of 26 patients (25%) were excluded from analysis secondary to poor quality of the data. In 21 of these 26 patients (81%) the poor quality data was obtained during the portable device segment of the study. OSA was defined as a respiratory disturbance index (RDI) $\geq$ 15. Based on calculations from the information provided, the sensitivity and specificity for the portable monitoring study were 81% and 98% respectively.

Orr (1994) studied 40 patients, 20 from each of 2 sleep laboratories with portable and laboratory based studies performed simultaneously in the laboratory setting. Sleep I/T, the portable device utilized, was described as an 8 channel device. No results had to be excluded from the analysis secondary to missing or unanalyzable data. Sleep I/T data were analyzed automatically. Based on an RDI $\geq$ 15, the sensitivity and specificity for the Sleep I/T system were 100% and 93% respectively.

Mykytyn (1999) studied 20 male symptomatic patients referred to a sleep laboratory for suspected OSA. Portable and laboratory based studies were performed simultaneously in the laboratory setting. For the portable study, patients were randomly assigned to either an attended or unattended group. Compumedics PS1-Series was the portable device utilized. Outcome measures included determination of AHI, signal quality, derived values such as sleep staging and efficiency, and clinical interpretation of the data by an experienced sleep physician. Using an AHI > 10 as diagnostic of OSA, the sensitivity and specificity of the portable device were 80% and 90% respectively. For the diagnostic cut-off of AHI > 20, the sensitivity and specificity of the portable device were reported as 100% and 100% respectively. Based on the physician's interpretation of the data, diagnostic concordance was attained in 16 of 18 study pairs (89%). Two portable studies, one attended and one unattended, were deemed insufficient for analysis secondary to poor quality data. The diagnostic interpretation for the 2 discordant pairs were diagnosing OSA versus upper airway resistance and diagnosing moderate versus mild OSA.

Two other studies, Iber (2004) and Fry (1998), also compared unattended portable devices with laboratory based PSG. No sensitivity or specificity data were provided. The primary outcome measures included quality of the recordings, reliability of neurophysiologic and cardiorespiratory parameters of sleep, and measurement of RDI obtained using the portable device versus the laboratory based PSG. Iber (2004) reported results as interclass correlational data based on the reproducibility of measurements for RDI when comparing unattended portable studies with laboratory based studies. Twelve participants were excluded from the analysis secondary to poor quality data of either the unattended or laboratory based study. Correlational data showed reproducibility of RDI measurements obtained using the portable device when compared to the laboratory study. Fry (1998) reported results using the Pearson correlation coefficient. All data were interpretable. Correlational data showed a moderate to strong degree of agreement for sleep and respiratory parameters obtained with the portable study when compared to the laboratory study, r values were in the range of 0.775 - 0.999.

For portable device studies performed unattended in the home by Portier (2004) and Fry (1998), the authors noted that patients came into the laboratory for education on proper use of the device. In addition, they received assistance with device set-up which could include proper application of the sensors in the Fry study. Portier (2004) also noted that the devices were tested to make sure they were functioning properly. Iber (2004) noted that patients had electrodes attached immediately before sleep. The authors did not provide further information regarding patient education or technician assistance.

Sixteen studies were reviewed that addressed the diagnostic test performance of portable multi-channel home sleep testing devices with a minimum of four monitored channels including: ventilation or airflow (at least two channels of respiratory movement or respiratory movement and airflow), heart rate or ECG, and oxygen saturation. Laboratory based polysomnography was used as the reference standard.

Three studies[14] evaluated the use of unattended portable monitoring devices in the home setting and compared the results to laboratory based polysomnography. Ancoli-Israel (1997) studied 36 volunteer subjects recruited from a larger study. Patients first underwent an in-home portable study followed by two nights of laboratory polysomnography within one week of the initial study. Patients received in laboratory set-up of the portable device, the Nightwatch System. Data were scored automatically with the ability to manually verify the information. Thirty-four subjects had data available for analysis. Two subjects did not have analyzable dataone during the portable device study and one during laboratory polysomnography. Based on an AHI $\geq$ 10, the sensitivity and specificity for the portable device were 100% and 63% respectively.

Parra (1997) performed a study of 89 patients referred to a sleep clinic for evaluation of OSA. Within a one-month period, patients underwent both laboratory and home based studies. Fifty of 89 patients had technician assistance in setting up the equipment in their home. The EdenTrace system was the portable device utilized in this study. Primary outcome measures were diagnostic agreement, determination of diagnostic usefulness, and clinical decision making. Using the Bland and Altman method for determining diagnostic agreement, agreement was noted in AHI measurements obtained by both study methods. The sensitivity and specificity for the portable device were calculated for various AHI cut-off points10, 18, and 23. Based on polysomnographic AHI > 10 and portable device AHI > 18 and diagnostic of OSA, the sensitivity and specificity for the portable device were 73% and 80% respectively. Based on the same polysomnographic cut-off point and portable device AHI > 23, the sensitivity and specificity for the portable device were 63% and 93% respectively. When comparing portable and laboratory based studies, clinical decision making was the same for 79 (89%) of patients. Of the 10 patients with discordant results, six would not have received CPAP therapy based on the portable study but would have received it based on polysomnography.

Whittle (1997) performed a study of patients referred to a sleep clinic for suspected OSA. The two part study consisted of both a validation and prospective trial. The EdenTrace system, the portable device used, was a four channel device. Twenty-three subjects underwent the validation study which included laboratory based polysomnography on the first night, followed by an unattended home study on the second night. Twenty of 23 studies (87%) produced interpretable recordings and were used in data analysis. A significant correlation (r = 0.8) was found when comparing AHI obtained using the laboratory and home based studies. Based on the validation study, an AHI > 30 was chosen as diagnostic of OSA for home based studies. One hundred and forty-nine subjects took part in the prospective trial. Twenty seven of 149 home based studies (18%) were uninterpretable. Patients with an AHI of < 30 based on the home studies and symptoms of daytime sleepiness were further investigated with laboratory based polysomnography. Fifty eight subjects had data from both studies that could be used for comparison. The sensitivity and specificity for the home based study based on an AHI > 30 for the home study and AHI > 15 for polysomnography were 75% and 58% respectively. The authors also performed a control arm of the study that included 75 patients referred to the sleep clinic who only received laboratory polysomnography.

Four studies[15] compared the use of unattended portable monitoring devices in the home setting to simultaneous polysomnography and portable monitoring performed in the laboratory setting. Dingli (2003) studied 101 patients referred to a sleep clinic for sleep related complaints. Patients were assigned to two groups. Forty underwent synchronous polysomnography and portable studies in the laboratory. The remaining 61 patients received an in-home unattended study and in-lab polysomnography on separate nights. Patients did receive technician instructions on how to operate the equipment in the sleep clinic prior to taking the device home. The portable device utilized was the Embletta system. Results for the synchronous study excluded one patient because no data was recorded on the Embletta system. Eleven of 61 (18%) of home study patients were excluded from analysis secondary to inadequate recordings. Based on polysomnography scoring AHI $\geq$ 15 as diagnostic of OSA and Embletta scoring of (A+H) x hrs in bed $\geq$ 20, the Embletta system had an accuracy of 100% (23/23) for identifying persons with disease. Based on the Embletta system, nine patients were classified as not having OSA with (A+ H) x hrs in bed $\leq$ 10 and all of these patients had polysomnography AHI $\leq$ 15. Therefore, the diagnostic accuracy for determining persons without disease was 100% (9/9). Eighteen patients (36%) were classified as possibly having OSA based on the Embletta system and (A+H) x hrs in bed $\geq$ 10 but $\leq$ 20 and would likely have required additional testing for definitive diagnosis. Fifteen of these 18 patients would have been diagnosed as having OSA based on polysomnography. The sensitivity and specificity were not explicitly stated for the Embletta system but were calculated as 61% and 75% respectively.

Reichert (2003) studied 51 patients referred to a sleep laboratory for clinical suspicion of OSA. Patients underwent simultaneous polysomnography and attended portable device studies in the laboratory. Patients also underwent 3 nights of separate unattended home recordings and the average AHI across the 3 nights of studies were used for results. Forty-five patients had data that were used in the analysis. AHI $\geq$ 15 was used as diagnostic of OSA for both polysomnography and the portable device. Six percent (3/48) of patients did not have interpretable data secondary to problems with the portable monitoring device in the home setting. Thirteen percent (7/51) of patients did not have data from the portable monitoring device in the laboratory setting due to technician error or data loss. The sensitivity and specificity for the attended portable device studies were 95 ± 5% and 91 ± 6% respectively. The sensitivity and specificity for the unattended portable device studies in the home setting were 91 ± 6% and 83 ± 8% respectively. The authors also compared sensitivity and specificities for the attended and unattended use of the device. The sensitivities for attended in lab use and unattended at home use were 94 ± 5% and 89 ± 7% respectively. The specificities for attended in lab use and unattended at home use were 90 ± 6.7% and 80 ± 8.9%.

White (1995) studied 100 patients referred for evaluation of sleep related complaints. Thirty patients underwent simultaneous portable studies and polysomnography in the laboratory setting. Seventy patients underwent laboratory based polysomnography and an additional portable study in the home setting. The Nightwatch System, the portable device utilized, has the capability to transmit signals in real time to the sleep laboratory. Throughout the night, if problems with the signal were identified by the lab technician, patients were telephoned and instructed on how to correct the problem. Two (2.8%) home studies were excluded from analysis secondary to lack of interpretable data. Eighty-one percent (57/70) of participants required a phone call by the technician to correct equipment or signal problems. The sensitivity and specificity of the portable device when used in the lab were provided for two different AHI cut-off points. For AHI > 10, the sensitivity and specificity were 100% and 64% respectively. Positive and negative predictive values were 87% and 100% respectively. For AHI > 20, the sensitivity and specificity were 77% and 88% respectively. Positive and negative predictive values were 83% and 83% respectively. The sensitivity and specificity for the portable device when used at home for AHI > 10 were 91% and 71% respectively. Positive and negative predictive values were 87% and 84% respectively. For AHI > 20, the sensitivity and specificity were 86% and 83% respectively. Positive and negative predictive values were 79% and 89% respectively.

Redline (1991) studied 51 subjects including a mix of healthy volunteers, relatives of apnea patients, persons with sleep related complaints, and patients with pulmonary disorders. Results were reported for 20 subjects who underwent simultaneous portable studies and laboratory based polysomnography and 5 subjects who underwent separate unattended portable studies and laboratory based polysomnography. RDI was noted to be highly correlated (r = 0.96) when comparing devices. Using an RDI $\geq$ 10 by PSG as diagnostic of OSA, 95% (20/21) patients would have been accurately diagnosed by the portable study.

Studies performed by Parra (1997), Dingli (2003), Reichert (2003), and White (1995) detailed the methodology used for unattended portable device set-up. Fifty of 89 patients in the Parra study had the technologist set-up the equipment. The remainder of the patients was provided written instructions and 10 minutes of technician instruction. Dingli (2003) also provided education to patients in the form of written instructions. However, patients were responsible for unsupervised equipment setting in their homes. In the Reichert study patients were given written instruction and no other form of assistance with device set-up. The device used included a voice alert system that would awaken patients and alert them if any of the sensors became dislodged during the night. In the study by White, all of the patients came into the laboratory for instructions on proper use of the device. The majority also had the device hooked-up while they were in the laboratory. This device also transmitted signal information back to the laboratory in real-time. At the start of the study, the laboratory could confirm that the device was working properly. In addition, throughout the course of the night the technician could contact the patient to correct any signaling or equipment problems.

Nine additional studies[16] evaluated the results of various portable monitoring devices used simultaneously with laboratory based polysomnography. Each study involved the use of various endpoints including Apnea Index (AI), Apnea Hypopnea Index (AHI), and/or Respiratory Distress Index (RDI). Sample sizes ranged from 29 to 150 participants. For these studies the average participant age was 52. Most studies involved the use of consecutive patients referred to sleep lab for evaluation. Only two studies, Claman (2001) and Verse (1998), included information about inclusion and exclusion criteria.

In all nine studies, type III devices and PSG were performed simultaneously in the laboratory setting and an attendant was present for all studies. No studies were performed in the home setting. In the study by Calleja (2002), the unattended mode was selected for use in portable monitoring device. All other studies, when dealing with type III devices, made no mention of operating in an unattended mode. In all of the studies the technician placed the sensors for both the type III devices and PSG. One study noted that technicians controlled PSG recordings, and were allowed to fix any failing signals, (Marrone 2002). In this same study, technicians were not allowed to visualize signal recordings from type III devices. If during the study one of the polysomnography sensors malfunctioned or became detached, the technician would correct it. The other studies did not describe how this same problem is addressed for type III devices.

Several of the nine studies used measures of agreement such as correlation coefficients or the Bland and Altman analysis to determine the degree of association between endpoints for portable monitoring devices as compared to PSG. Six of the nine studies used Pearson's correlation coefficients to determine the degree of association between endpoints for both diagnostic modalities. AHI values varied between studies; some defined OSA with an AHI of 10, while other studies used an AHI of 15. Man (1995) reported a correlation coefficient of 97% for AHI between both diagnostic tests, while Claman (2001) reported a correlation of 96% when comparing AHI as the endpoint. Verse (2000), noted a correlation coefficient of 97% when using AI as an endpoint for comparison, while Marrone (2001) noted significant correlation when comparing a number of indices between both diagnostic tests (r between 68% and 99%). Esnaola (1996) noted an intraclass correlation agreement of 72% for AHI between the two diagnostic tests. Another statistical test used to measure the degree of agreement between diagnostic tests was the Bland and Altman analysis. Both the Marrone (2001) and the Ballester (1995) studies showed high levels of agreement for indices between diagnostic tests.

Sensitivity and specificity were used by all studies to measure accuracy between both diagnostic procedures (refer to Appendix A). Positive predictive values (PPV) and negative predictive values (NPV) were also reported in some studies. A few studies employed the use of receiver operating characteristic (ROC) curves to determine these measures of accuracy Ballester (1995); Esnaola (1996); and Calleja (2002). Most studies used AHI threshold values ranging from 5 to 30, though two studies; Ficker (2001) and Zucconi (1996); compared AHI values as high as 40. A large number of studies had dichotomous endpoints (e.g., AHI values < 15, or AHI values > 15). One study, Calleja (2002), included five different sets of ranges for AHI values.

When reviewing the studies using dichotomous outcomes, one study Claman (2001) reported sensitivity for AHI > 15, but did not report specificity for this same variable. It also reported specificity for AHI < 15, but did not report sensitivity for this variable. Another dichotomous study, Ballester (2000), developed a receiver operating characteristic (ROC) curve to predict accuracy measures using PSG cut-off values. When reviewing these studies with dichotomous values, it is noted that all studies reveal high values for sensitivities as well as specificities. Studies also show high positive predictive values as well as negative predictive values. In general for these studies with dichotomous values, as we move from a lower to a higher AHI value, the sensitivity for this variable either stays the same or increases in value thus indicating a strong accuracy measure compared to PSG in making a diagnosis of obstructive sleep apnea. Also, the corresponding specificity increases in value. This also indicates that agreement exists with PSG in excluding a diagnosis of obstructive sleep apnea.

A number of other studies were evaluated comparing PSG with level III devices, using three or more sets of AHI ranges. Two studies, Esnaola (1996) and Zucconi (1996), compared results of manual scoring of AHI to automatic scoring in its comparison to PSG. Both studies consistently demonstrated that manual scoring was superior to automatic scoring. The Esnaola study used changes in heart rate, oxygen saturation, and breathing sounds as indices to identify occurrences of apnea or hypopnea during manual scoring. The study utilized a system of two or three channel manual scoring indices. In the case of the three-channel manual scoring index (MS3), an event was defined as the simultaneous occurrence of changes in all three variables. The two-channel manual scoring index (MS2) defined an event when changes in two of the three variables were identified. Specificity was high for both two and three-channel manual scoring systems, but sensitivity was marginal, especially for the two-channel system. This study also revealed that sensitivity as well as negative predictive value was increased when measuring with a two-channel system as compared to the three-channel system. There is a slight increased positive predictive value for the two-channel system as compared to the three-channel system.

The study performed by Zucconi (1996) revealed that even over a large range of AHI scores (10 through 40), all measures of accuracy (sensitivity, specificity, PPV, NPV) using the manual scoring were consistently high indicating close agreement with PSG. Automatic scoring was in agreement with PSG reading until AHI values were high (e.g., AHI > 40). At this level the accuracy in comparison with PSG became poor.

Three studies were also of particular interest. In both the Ficker (2001) study and the Verse (2000) study, as AHI values increased from 5 to 40, the specificities remained consistently high, while the sensitivities decreased. However, in the Calleja (2002) study, as AHI values increased from 5 to 30, the sensitivities remained essentially high. The specificities were initially high but then the values started to fall until AHI of 15. After that point, the specificities began to rise again. This seems to indicate that the device was most accurate at extreme AHI values (5 and 30), but was less accurate for intermediate values (10 to 20).

Three studies provided information on loss of data. In one study, Zucconi (1996), the data from one patient was not considered in the investigation due to loss of data. In a second study, Calleja (2002), eight percent of sleep studies were invalid due to sleep times less than 240 minutes, lack of thermistor signal, or incomplete recording due to technical problems. One final study, Esnaola (1996) recruited 152 participants, but due to recording problems, was only able to report the results of 150 subjects. None of the other studies noted attrition due to loss of data.

4. MCAC

On November 24, 1998, the Secretary of the Department of Health and Human Services chartered the Medicare Coverage Advisory Committee (MCAC). The MCAC advises CMS on whether specific medical items and services are reasonable and necessary under Medicare law. They perform this task via a careful review and discussion of specific clinical and scientific issues in an open and public forum. The MCAC is advisory in nature, with the final decision on all issues resting with CMS. Accordingly, the advice rendered by the MCAC is most useful when it results from a process of full scientific inquiry and thoughtful discussion, in an open forum, with careful framing of recommendations and clear identification of the basis of those recommendations. The charter was renewed on November 22, 2002.

The MCAC is used to supplement CMS's internal expertise and to ensure an unbiased and contemporary consideration of "state of the art" technology and science. Accordingly, MCAC members are valued for their background, education, and expertise in a wide variety of scientific, clinical, and other related fields. In composing the MCAC, CMS was diligent in pursuing ethnic, gender, geographic, and other diverse views, and to carefully screen each member to determine potential conflicts of interest. All MCAC members are trained, appointed, and then perform their service as members of an MCAC panel.

The MCAC met on September 28, 2004, to discuss and make recommendations to CMS concerning the quality of the evidence and related issues for the use of portable multichannel home sleep testing as an alternative to facility based polysomnography. The MCAC transcript, minutes, and presentation are available at: https://www.cms.hhs.gov/mcd/viewmcac.asp?id=110

Refer to Appendix C for the MCAC questions and scoring summary. The possible scores ranged from 1 (low) to 5 (high). Generally, the panel votes were low for the majority of questions presented. Overall, the panel members were moderately confident that multi-channel home sleep study testing would improve patients' health outcomes and are as accessible as facility-based polysomnography in the diagnosis of OSA.

5. Evidence-based guidelines

Diagnosis and Treatment of Obstructive Sleep Apnea (2004): Institute for Clinical Systems Improvement

The guidelines recommend the use of unattended portable recording device studies for patients with a high pretest probability of obstructive sleep apnea/hypopnea syndrome (OSAHS) as an acceptable alternative to standard polysomnography in certain situations: (1) patients with severe clinical symptoms that are indicative of a diagnosis of obstructive sleep apnea and when initiation of treatment is urgent and standard polysomnography is not readily available, (2) patients unable to be studied in the sleep laboratory, and (3) follow-up studies when diagnosis has been established by standard polysomnography and therapy has been initiated.

Management of obstructive sleep apnoea/hypopnoea syndrome in adults (2003). A national clinical guideline. Scottish Intercollegiate Guidelines Network (SIGN).

Limited sleep studies to assess respiratory events are an adequate first-line method of diagnostic assessment for obstructive sleep apnoea/hypopnoea syndrome (OSAHS).

6. Professional Society Position Statements

A search for published professional society position statements on the use of home sleep monitoring for OSA yielded the following results.

The American Thoracic Society (ATS), the American College of Chest Physicians (ACCP), and the American Academy of Sleep Medicine (AASM) cosponsored a working group and hired an evidence-based practice center to produce a detailed literature search and evidence review on the use of portable monitors for investigating patients with suspected sleep apnea. (*Chest.* 2003;124:1543-1579.)

The use of portable monitoring as an initial diagnostic tool in selected patients may reduce costs by lowering the use of resources and allowing patients to proceed directly to CPAP titration studies if the test results were positive, and in some cases to forego additional testing if the test results were negative. The limited generalizability of these studies warrants caution since the conclusions were heavily dependent on the pretest probability and the threshold level for the diagnosis of sleep apnea. Future studies are clearly needed to add further perspective, and should include formal cost-benefit analyses comparing portable monitoring to split-night polysomnogram protocols and assessing the ultimate result on patient outcomes with appropriate treatment follow-up. http://www.chestjournal.org/cgi/content/full/124/4/1543. Accessed Dec 17, 2004)

The American Academy of Pediatrics Clinical Practice Guideline: Diagnosis and Management of Childhood Obstructive Sleep Apnea Syndrome describes PSG as the "gold standard" in their algorithm for the diagnosis of OSA in children. (PEDIATRICS Vol. 109 No. 4 April 2002, pp. 704-712)

"Unattended home polysomnography in children has been evaluated by only 1 center.[48] Home polysomnography yielded similar results to laboratory studies. However, it should be noted that the equipment used in this study was relatively sophisticated and included respiratory inductive plethysmography (a method for determining ventilation without using oronasal sensors), oximeter pulse wave form, and videotaping. The utility of unattended home studies in children using commercially available 4- to 6-channel recording equipment has not been studied." http://aappolicy.aappublications.org/cgi/content/full/pediatrics;109/4/704. Accessed Dec 17, 2004)

Several professional societies presented position statements during the comment periods or at the MCAC.

The American Association for Homecare supports a revision to the current NCD to permit the use of portable multi -channel sleep testing devices in the home site of service as an alternative to facility based polysomnography for the evaluation of OSA.

The Association of Polysomnography Technologists stated that there are many "nonfacility" sites that do an excellent job at accurately diagnosing and treating OSA, and that excluding those sites will severely limit those Medicare recipients to potentially life saving care.

The National Association for Medical Direction of Respiratory Care acknowledged that the clinical literature regarding home testing is not supportive, but notes that many of these publications were based on outdated equipment. The association believes that timely access to sleep laboratories is a significant problem, and that a reimbursement structure that would permit home sleep studies under some circumstances, such as accreditation by an appropriate body, would address part of that access issue.

The American Academy of Otolaryngology-Head and Neck Surgery supports the use of multi-channel home sleep study testing as a potential cost-effective alternative to PSG and as means of improving access to care for the large adult population at risk for sleep apnea.

The Sleep Research Society expressed disapproval of the requested policy change based on currently available research findings.

The New England Polysomnographic Society expressed opposition to the proposed coverage of home polysomnography.

7. Expert Opinion

CMS met with the manufacturers and providers of portable home sleep monitors, and with physicians who are affiliated with facility-based and/or home based sleep study programs.

On May 13, 2004, representatives of Aircare Home Medical Company, met with CMS to support the modification of the current national coverage guidelines specifying that only a polysomnography done in a facility –based sleep laboratory may be used to identify and diagnose patients with obstructive sleep apnea.

On June 3, 2004, representatives of VioMetrics met with CMS on Medicare's coverage of home sleep testing. The group expressed their support for Dr. Davidson's request for Medicare coverage of multi-channel home sleep study testing. In addition, the group demonstrated one of the suggested new generation devices capable of doing a home sleep study test, the LifeShirt System.

On June 10, 2004, representatives of Oxford BioSignals, met with CMS to support the use of multi-channel home sleep study testing as an alternative to facility-based polysomnography.

On July 22, 2004, Dr. Terrence M. Davidson, M.D., the requestor of the NCD met with CMS to give a presentation to support the use of multi-channel home sleep study testing as an alternative to facility-based polysomnography.

On August 25, 2004, representatives of Snap Laboratories, met with CMS to support the NCD for multi-channel home sleep study testing. The representatives suggested that CMS develop a patient profile of which patient requires a PSG in a hospital sleep laboratory.

On August 31, 2004, CMS met with several sleep physicians to support Dr. Davidson's proposal that CMS cover type 3 monitors for the diagnosis of obstructive sleep apnea syndrome.

On September 16, 2004, representatives from Sleep Solution and St. Lukes Roosevelt Hospital met with CMS in support of the request for Medicare coverage of on multi-channel home sleep study testing.

8. Public Comments

During the first comment period CMS received 82 comments. We received very few comments from Medicare beneficiaries. The majority of the comments in favor of using multi-channel home sleep study testing as an alternative to a polysomnography in a facility-based sleep study laboratory came from medical institutions/centers, physicians, respiratory therapist, laboratories, industry representatives and associations. The main concern expressed by those in favor of the use of multi-channel home sleep study testing was the patient's lack of timely access to facility based polysomnography.

The commenters not in favor of CMS revising the current policy for CPAP represent medical institutions/sleep centers, physicians, respiratory therapists, and associations. These commenters do not believe that there is sufficient evidence available to support the use of home sleep study testing, and questioned whether the use of home sleep testing would actually reduce costs.

During the second public comment period, June 25-July 26, we received 35 comments on the following two questions:

Question a: If unattended portable multi-channel home sleep testing is as effective as polysomnography in the diagnosis of obstructive sleep apnea which parameters of sleep and cardiorespiratory function (i.e. sleep staging, body position, limb movements, respiratory effort, airflow, oxygen saturation, ECG) are required?

Some commenters believed that only oxygen saturation is needed, others thought that 3 parameters are needed, i.e. oximetry, nasal air flow and chest excursion. Others stated that the requirements for a home study should not be any different from a laboratory tests.

Question b: If unattended portable multi-channel home sleep testing is as effective as polysomnography in the diagnosis of obstructive sleep apnea, what conditions (i.e. patient education, technician support) are required so that it is done correctly in the home?

Some commenters believe that the individuals board certified in Sleep Medicine would be most likely to provide the most effective and comprehensive care to patients suspected of having sleep apnea. Other commenters thought that the patient must be trained by a respiratory therapist, a registered polysomnographic technologist or physician to use the device approximately.

VIII. CMS Analysis

National coverage determinations (NCDs) are determinations by the Secretary with respect to whether or not a particular item or service is covered nationally under title XVIII of the Social Security Act § 1869(f)(1)(B). In order to be covered by Medicare, an item or service must fall within one or more benefit categories, and must not be otherwise excluded from coverage. Moreover, with limited exceptions, the expenses incurred for items or services must be "reasonable and necessary for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member." § 1862(a)(1)(A).

This section represents the agency's evaluation of the evidence considered and tentative conclusions reached for the assessment questions.

Question 1: How does the diagnostic test performance of unattended portable multi-channel home sleep testing devices compare to facility-based polysomnography in the diagnosis of OSA?

a. If unattended portable multi-channel home sleep testing is as effective as polysomnography in the diagnosis of obstructive sleep apnea, which parameters of sleep and cardiorespiratory function (i.e. sleep staging, body position, limb movements, respiratory effort, airflow, oxygen saturation, electrocardiogram) are required?

b. If unattended portable multi-channel home sleep testing is as effective as polysomnography in the diagnosis of obstructive sleep apnea, what conditions (i.e. patient education, technician support) are required so that it is done correctly in the home?

Clinical Considerations
Sleep disordered breathing is a common disorder and is increasingly being recognized as having a significant impact on the health and quality of life of persons affected. Laboratory based polysomnography with continuous overnight monitoring of various neurophysiologic and cardiorespiratory parameters of sleep has been the mainstay in the diagnostic work-up for persons suspected of having OSA. A variety of portable devices have been developed that measure, to varying extents, the same or similar neurophysiologic and cardiorespiratory parameters of sleep in an effort to appropriately diagnose OSA and address issues such as accessibility, delay in diagnosis, and ease of diagnosis in a setting that would more closely match the patients routine sleeping habits. With the advent and use of these portable monitoring devices, a number of issues have arisen regarding their widespread use, including (1) ability to consistently and accurately diagnose OSA (2) determination of exactly which neurophysiologic and/or cardiorespiratory parameters of sleep are required for diagnosis; (3) determination of the appropriate cut-off point for diagnosis based on portable device scoring which may or may not be the same cut-off typically used for polysomnography; (4) loss of data and need for repeat testing with unattended use of portable devices; and (5) use of automated/computerized portable device scoring and diagnosis that does not permit an actual trained person to review the data. A portable monitoring device that could reliably and accurately rule-in and/or rule-out persons suspected of having OSA would be beneficial in aiding clinical decision making and has the potential to significantly impact net health outcomes.

Type II Devices
Only five studies reviewed addressed the diagnostic test performance of portable multi-channel devices with a minimum of 7 monitored channels (Type II Devices, based on the 1994 ASDA classification system). Sample sizes were relatively small, the largest being 103 study participants. In addition, the majority of subjects studied were men between the ages of 45-55. Three of the 5 studies (Portier (2000), Iber (2004), and Fry (1998)) performed unattended portable studies in the home setting and compared the results to laboratory based polysomnography. Only the study by Portier (2000) provided sensitivity and specificity data. While the test was specific, identifying 98% of persons without disease, it had a poor sensitivity. It only correctly identified 81% of persons with disease. In addition, when comparing exclusion of data from the analysis, 81% (21/26) of participants who had their results excluded from the analysis were those obtained during the unattended portable device study compared to 19% (5/26) during polysomnography.

The studies undertaken by Iber (2004) and Fry (1999) provided results in the form of averages based on participant questionnaires and correlational data for sleep, respiratory, and limb parameters measured during both the unattended home study and laboratory based polysomnography. While this information reports high correlation between measurements obtained with both devices, correlational data are generally a representation of the strength or relatedness of two measures. They are surrogate markers for how well one diagnostic test will accurately perform in relation to another for appropriate diagnosis of disease. Measures such as sensitivity, specificity, positive or negative predictive values, and likelihood ratios may be better suited in determining the utility of a diagnostic test in a particular patient population[17]

The studies performed by Mykytyn (1999) and Orr (1994) measured the performance of the portable device under ideal conditions in the laboratory. This calls into question their generalizability to the unattended home setting. Mykytyn included a subset of patients (10) who had unattended studies in the laboratory setting. In addition, the authors compared AHI cut-off points of 10 and 20 and calculated sensitivities and specificities for each. The sensitivity for the lower AHI was reported as 80%, while the higher AHI had a reported sensitivity of 100%. The study did not provide comparative sensitivity and specificity data for unattended study results.

Potential impact of the use of Type II portable monitoring devices on net health outcomes.
Only one study, Mykytyn (1999), included an analysis of clinical decision making based on the results of polysomnography and a portable monitoring device. Diagnostic concordance between the physician's interpretation of the data and results obtained by polysomnography or portable device monitoring was achieved in 89% of cases. Difficulty in diagnosis surrounded two cases. The diagnostic dilemma surrounded distinguishing OSA from another disorder and distinguishing severity of OSA. As previously stated, this study only had 20 participants and the generalizability of these results to the general or Medicare populations is unclear.

Type III Devices
Seventeen studies reviewed addressed the diagnostic test performance of portable multi-channel devices with a minimum of four monitored channels (Type III Devices, based on the 1994 ASDA classification system). Since type III devices do not include EEG monitoring, they are not able to calculate total sleep time. Based on this, though AHI is an endpoint that is being investigated, its definition, as well as its components may differ between the two diagnostic tests. Therefore, it can be said that the two different tests might not be measuring the same endpoint. This problem was common in all the studies evaluated. Eight of the seventeen studies compared the use of unattended portable devices to laboratory based polysomnography. Sample sizes were also relatively small with the largest study utilizing data from 149 participants. Three studies[18] utilized two different cut-off points for AHI as diagnostic of OSA. One cut-off was chosen for the portable device and another for polysomnography. Use of two different cut-off points suggests that determining the appropriate number of AHIs for diagnosis using the portable device is a somewhat imprecise process. With the use of one AHI for the portable device as corresponding to a different AHI for polysomnography, the sensitivities for those three studies was in the range of 61-75% and specificities in the range of 58-75%. Findings from these studies suggest the need for further standardization of the portable devices and determination of an appropriate AHI cut-off point that would reliably and accurately rule-in or rule-out disease.

Three studies[19] used the same AHI cut-off point for both the portable device and laboratory based polysomnography. When comparing AHI cut-off points between 10-15, the sensitivities for the portable devices in these studies was in the range of 95-100% and specificities in the range of 63-91%. One study, Reichert (2003), even compared a small subset of patients to evaluate differences in sensitivity and specificity for attended and unattended use of the portable device. The sensitivities and specificities for this comparison appeared to overlap. Findings from these three studies suggest that these portable devices may have similar operating characteristics to polysomnography. In addition, two studies[20] provided data to evaluate the accuracy of portable monitoring devices. For accuracy of ruling in disease, one study reported 95% accuracy while the other reported 100% accuracy. For accuracy in ruling out disease, only one study reported results and they were 100% but only correspond to data on nine patients. These findings suggest that for persons in whom the clinical suspicion of disease is high, portable monitoring devices may assist in clinical decision making. While the findings for ruling out disease are hopeful, the results are tempered by the small number of patients.

In the nine studies that evaluated type III devices and compared their accuracy with simultaneously performed polysomnography (PSG) in the diagnosis of OSA, a number of methodological flaws may call their conclusions into question. Since type III devices were tested simultaneously with PSG on participants, randomization is not an issue (each patient acted as his/her own control). Another problem that was noted in all studies, with the exception of the one performed by Claman (2001), is the lack of inclusion as well as exclusion criteria. The simultaneous examinations were performed in a referral center. No recordings using the type III device were performed in a home setting. The findings of these studies may not be applicable to a home setting. Another common problem found in these studies is an inconsistent definition of obstructive sleep apnea. Some studies used an AHI value of five to be diagnostic, while others used values of 10. Because of these differing definitions, patient selection could be affected, thus confounding the accuracy of both devices. Similarly, there existed a vast range of AHI values used as cut-offs. Some studies were very parsimonious in AHI ranges, while other studies employed large ranges. Due to the large range of AHI values used to define OSA, comparability between studies could be compromised.

The number of participants per study was relatively low (participants per study ranged from 29 to 150). More than half of the studies evaluating type III devices had less than 60 participants. Only one study acknowledged that data was not normally distributed and Wilcoxon's signed-rank test was used for comparison purposes Zucconi (1996). Also, as previously mentioned, the average age of study participant was 51. Findings in this age group may not necessarily be generalizable to the Medicare population.

In looking at individual studies, the study by Claman (2001) had a number of problems. First the sample size was low (n= 42). Also a consecutive sample of 42 patient volunteers was recruited for the study, which could result in selection bias. Also in this study, for PSG, the AHI was determined based on sleep time. For the Type III device, AHI was determined based on the total duration of recorded data. This is an issue common to many of the studies. Due to the use of different measures to determine AHI, there is a threat to internal validity, thus compromising the study. In the population studied, the mean ages of participants were 54. Generalizing these results to the Medicare population may be difficult. Sensitivities and specificities were also reported using PSG as a gold standard. But this study only reported sensitivities for patients with AHI > 15, and did not report specificities for this group. Likewise it reported specificities for patients with AHI < 15, but did not report the sensitivities. Though this study did have a number of flaws, it was only one of two studies which had inclusion as well as exclusion criteria (the study by Verse (2000) also had inclusion/exclusion criteria).

The study by Zucconi (1996) also suffered from a low participation (n= 29) rate. This study also used consecutive patients which were referred to a sleep study center, but the authors did not specify the source of the referral. Thus patient selection bias could skew the results. Another potential problem in this study was the lack of well-defined cut-offs for AHI as well as RDI. This could make assessment of validity of the device as difficult as the PSG interpretation since it could affect accuracy. To avoid this potential problem, the study adopted a test of diagnostic agreement which was previously applied by White (1995). But there has been no further validation of this agreement. The study also acknowledged that a small number of subjects (n = 8) with AHIs between 10 and 40 were recruited. This could result in a greatest potential for confusion of the diagnosis. Also, another limitation noted by the study was its choice of habitual snorers and suspected OSA patients, instead of the general population. This would make it difficult to generalize to the Medicare population.

The study by Man (1995) demonstrates some of the potential pitfalls between both diagnostic tests. According to the authors of the article, technicians took 45 minutes to 1½ hour to screen and validate the data used in type III devices. For the PSG, technicians typically spent 3 hours going over the records. Because of the increased time spent reviewing PSG data, potential bias against PSGs could exist due to increased scrutiny of PSG data and less scrutiny of type III recordings. There were also four occasions when the technician had to go into the sleep room to readjust sensors on the type III device. This might represent "lost data" if studies were performed in unattended settings. This study also endorsed the use of a Bland and Altman statistical analysis to assess agreement between diagnostic tests, using differences between PSG + type III device plotted against mean [(PSG + Poly G)/2] for both AI and AHI. In the final analysis of this test, it noted that there was close agreement, but did not provide the basis for this conclusion.

In determining the accuracy of type III devices, the study by Ballester (2000) employed the use of a logistic regression model to estimate the chances per unit of RDI of apnea. Receiver operating characteristic (ROC) curves were drawn to obtain sensitivity/specificity profile for RDI values obtained. In this study PSG cut-off points of < 10 and < 30 were chosen a priori for the purpose of diagnostic screening and for grading severity. Other studies have chosen AHIs as low as five and as high as 40 in their evaluation. If these same numbers had been evaluated, results might have shown that using this approach might result in less accuracy.

The study by Verse (2000) included 53 participants. This study also noted discrepancies in definitions of endpoints (AHI and AI). When evaluating the total group of patients, the values for AHI and AI measured by type III devices were lower than those noted in PSG. It further noted that the higher the value for AHI or AI using PSG, the greater the degree to which the level III devices underestimated the results. This discrepancy was probably due to the differences in measurement scope. Whereas the type III device based its index scores on the total measurement time (or total time in bed [TIB]), the corresponding indexes scores are calculated by PSG on the basis of total sleeping time (TST). Depending on which definition is used, a certain number of patients affected by sleep apnea could escape detection using a level III device. This discrepancy in measurement scope is a threat to the internal validity of the study.

The study by Esnaola (1996) used 152 consecutive patients suspected of obstructive sleep apnea in the comparison study between type III devices and PSG. As mentioned earlier, using consecutive patients for a study has a potential for selection bias. This study was limited to patients with suspected obstructive sleep apnea, so the results observed do not allow assessment of the diagnostic accuracy of type III devices when applied to subjects with other characteristics (e.g., asymptomatic patients), or in other settings (outside the sleep lab). This study also used ROC curves in determining the discriminatory ability of type III devices. Validity indices obtained from each of the ROC curves were estimated at cut-off points using two criteria, one as an exclusion test and the other as a confirmatory test. But because probability distributions of the cut-off points were not known, the standard errors of the selected cut-off points were not available. Furthermore the choice of the optimal cut-off point based on the values of the study participants tended to overestimate the validity indices (this could be especially important when dealing with small sample sizes). The authors of the study used the bootstrap method to estimate the optimal cut-off point, variability, bias and standard error of the corresponding validity indices (the bootstrap method is a method for estimating the sampling distribution of an estimator by resampling with replacement from the original sample). This latter maneuver seemed to complicate the analysis process even further.

The study by Calleja (2002) involved 79 patients (89% males) and also employed the use of ROC curves to determine the discriminatory ability of type III devices in making the diagnosis of obstructive sleep apnea. But unlike the study previously mentioned, there is no mention that exclusion or confirmatory testing was performed, nor any mention about using bootstrap method to estimate the standard error of the optimal cut-off point, and the bias and standard error of the corresponding validity indices. Because of the small sample size, preponderance of male participants, and relative young age of the subjects (mean age of 52), it is difficult to generalize the results to the Medicare population, especially to the non-referral segment.

The final study in the analysis of type III devices, Zucconi (1996), only had 29 consecutive participants. The use of consecutive participants makes the study prone to selection bias. One other weakness of this study is the use of habitual snorers and patients suspected of suffering from obstructive sleep apnea as sample subjects, instead of the general population. The study acknowledges that if it had used this approach, results would probably improve.

Potential impact of the use of Type III portable monitoring devices on net health outcomes.
One study, Parra (1997), provided an analysis of clinical decision making comparing the results of polysomnography to a portable monitoring device. Eighty-nine percent of the patients would have had the same diagnosis based on results of either test and been treated accordingly. There would have been a discrepancy in treatment with CPAP therapy in ten patients. These discrepancies were not based on the calculated AHI by either method, as patients still would have been identified accurately as either having or not having disease. However, treatment plans would have differed based on the severity of symptoms. Therefore, based on these results, portable devices provided accurate diagnosis and aided in clinical decision making for most patients. However, when symptom severity or other subjective measures where called into question, the results of the portable monitoring study may not be sufficient to aid in a more definitive diagnosis.

Summary

Question 1: How does the diagnostic test performance of unattended portable multi-channel home sleep testing devices compare to facility-based polysomnography in the diagnosis of OSA?

A small number of studies (five) addressed the diagnostic test performance of Type II devices as compared to facility-based polysomnography in the diagnosis of OSA. Only one of these studies, Portier (2000), reported sensitivity and specificity data for unattended portable studies performed in the home. The test accurately ruled out those persons without disease but was less accurate in correctly ruling in persons with disease. Another point of interest is evidenced by the fact that two studies, Portier (2000) and Orr (1994), used the same RDI value as diagnostic of OSA. While the specificities were comparable, the study performed in the unattended home setting had a 20% lower sensitivity than the study performed in the attended laboratory setting. As we have noted elsewhere in this draft decision memo, our confidence in the clinical application of sensitivity and specificity data to this particular diagnostic paradigm is tempered by the absence of a definitive gold standard to confirm true disease.

Seven of the sixteen studies that addressed the diagnostic test performance of Type III devices were performed unattended in the home. The ranges of sensitivities were 61-100%. The ranges of specificities were 58-83%. In addition, several studies used one value for the AHI cut-off for the portable device as diagnostic of OSA and a different value for the AHI cut-off for PSG. The use of two different cut-off points suggests the need for further standardization of portable devices.

a. If unattended portable multi-channel home sleep testing is as effective as polysomnography, which parameters of sleep and cardiorespiratory function (i.e. sleep staging, body position, limb movements, respiratory effort, airflow, oxygen saturation, electrocardiogram) are required?

Based on the studies reviewed, there is insufficient evidence to find that unattended portable multi-channel sleep testing devices are as effective as polysomnography.

b. If unattended portable multi-channel home sleep testing is as effective as polysomnography in the diagnosis of obstructive sleep apnea, what conditions (i.e. patient education, technician support) are required so that it is done correctly in the home?

Based on the studies reviewed, there is insufficient evidence to find that unattended portable multi-channel sleep testing devices are as effective as polysomnography in diagnosing obstructive sleep apnea. However, several studies did provide information regarding set-up for the unattended home study. Several modalities were employed to assist the patient in use of the equipment, including: 1) the patient came into the lab for technician-assisted set-up, 2) the patient came into the lab for technician instruction and written information, 3) real-time transmittal of the sleep study was provided to the sleep laboratory and a technician could call the patient when signaling problems arose, or 4) the device included a voice alert system to make the patient aware of a signaling or equipment problem. Sensitivity and specificity results did not seem to vary greatly based on the modality used.

Although laboratory based polysomnography is considered the "gold standard" for the diagnosis of sleep disorders, this approach has been called into question by some experts in the field of sleep medicine. There is insufficient clinical evidence available to assess the validity of laboratory based polysomnography in the diagnosis of OSA in adults.[21] In addition, a negative polysomnogram does not conclusively exclude the diagnosis of OSA for persons who present with a high clinical suspicion of disease. Alternative models have also been evaluated that use physical examination measures and other predictors such as body mass index and neck circumference to diagnosis disease and determine clinical decision making.[22] It is possible that for persons with a high pre-test probability of disease, a simple trial of CPAP therapy without prior diagnostic evaluation with polysomnographic studies may produce the same net health outcomes. Further research is needed that would not only address this issue but also more adequately evaluate issues raised when comparing portable devices to laboratory based polysomnography. These issues include addressing appropriate device set-up and minimizing data loss due to improper patient set-up or sensor failure, variability in the cut-off number chosen for AHI, scoring technique (manual vs. automatic), effect of portable device studies on clinical decision making and net health outcomes.

IX. Tentative Conclusion

The Centers for Medicare & Medicaid Services (CMS) is seeking public comment on the following proposed conclusions:

The evidence is not adequate to conclude that the use of unattended portable multi-channel sleep testing with a minimum of 7 monitored channels including EEG, EOG, EMG, ECG or heart rate, airflow, respiratory effort, and oxygen saturation (Type II Devices based on the 1994 ASDA classification) is reasonable and necessary in the diagnosis of OSA and these tests will remain noncovered for this purpose.

The evidence is not adequate to conclude that the use of unattended portable multi-channel sleep testing with a minimum of 4 monitored channels including ventilation or airflow, heart rate or ECG, and oxygen saturation (Type III Devices based on the 1994 ASDA classification system) is reasonable and necessary in the diagnosis of OSA and these tests will remain noncovered for this purpose

CMS is requesting public comments on this proposed decision memorandum pursuant to Section 731 of the Medicare Modernization Act. After considering the public comments and any additional evidence, we will make a final determination and issue a final coverage decision memorandum

Appendices

Appendix A- Evidence Tables [PDF, 315KB]

APPENDIX B

General Methodological Principles of Study Design
(Section VI of the Proposed Decision Memorandum)

When making national coverage determinations, CMS evaluates relevant clinical evidence to determine whether or not the evidence is of sufficient quality to support a finding that an item or service is reasonable and necessary. The overall objective for the critical appraisal of the evidence is to determine to what degree we are confident that: 1) the specific assessment questions can be answered conclusively; and 2) the intervention will improve net health outcomes for patients.

We divide the assessment of clinical evidence into three stages: 1) the quality of the individual studies; 2) the generalizability of findings from individual studies to the Medicare population; and 3) overarching conclusions that can be drawn from the body of the evidence on the direction and magnitude of the intervention's potential risks and benefits.

The methodological principles described below represent a broad discussion of the issues we consider when reviewing clinical evidence. However, it should be noted that each coverage determination has its unique methodological aspects.

Assessing Individual Studies

Methodologists have developed criteria to determine weaknesses and strengths of clinical research. Strength of evidence generally refers to: 1) the scientific validity underlying study findings regarding causal relationships between health care interventions and health outcomes; and 2) the reduction of bias. In general, some of the methodological attributes associated with stronger evidence include those listed below:

- Use of randomization (allocation of patients to either intervention or control group) in order to minimize bias.
- Use of contemporaneous control groups (rather than historical controls) in order to ensure comparability between the intervention and control groups.
- Prospective (rather than retrospective) studies to ensure a more thorough and systematical assessment of factors related to outcomes.
- Larger sample sizes in studies to demonstrate both statistically significant as well as clinically significant outcomes that can be extrapolated to the Medicare population. Sample size should be large enough to make chance an unlikely explanation for what was found.
- Masking (blinding) to ensure patients and investigators do not know to which group patients were assigned (intervention or control). This is important especially in subjective outcomes, such as pain or quality of life, where enthusiasm and psychological factors may lead to an improved perceived outcome by either the patient or assessor.

Regardless of whether the design of a study is a randomized controlled trial, a non-randomized controlled trial, a cohort study or a case-control study, the primary criterion for methodological strength or quality is the extent to which differences between intervention and control groups can be attributed to the intervention studied. This is known as internal validity. Various types of bias can undermine internal validity. These include:

- Different characteristics between patients participating and those theoretically eligible for study but not participating (selection bias).
- Co-interventions or provision of care apart from the intervention under evaluation (performance bias).
- Differential assessment of outcome (detection bias).
- Occurrence and reporting of patients who do not complete the study (attrition bias).

In principle, rankings of research design have been based on the ability of each study design category to minimize these biases. A randomized controlled trial minimizes systematic bias (in theory) by selecting a sample of participants from a particular population and allocating them randomly to the intervention and control groups. Thus, in general, randomized controlled studies have been typically assigned the greatest strength, followed by non-randomized clinical trials and controlled observational studies. The design, conduct and analysis of trials are important factors as well. For example, a well designed and conducted observational study with a large sample size may provide stronger evidence than a poorly designed and conducted randomized controlled trial with a small sample size. The following is a representative list of study designs (some of which have alternative names) ranked from most to least methodologically rigorous in their potential ability to minimize systematic bias:

- Randomized controlled trials
- Non-randomized controlled trials
- Prospective cohort studies
- Retrospective case control studies
- Cross-sectional studies
- Surveillance studies (e.g., using registries or surveys)
- Consecutive case series
- Single case reports

When there are merely associations but not causal relationships between a study's variables and outcomes, it is important not to draw causal inferences. Confounding refers to independent variables that systematically vary with the causal variable. This distorts measurement of the outcome of interest because its effect size is mixed with the effects of other extraneous factors. For observational, and in some cases randomized controlled trials, the method in which confounding factors are handled (either through stratification or appropriate statistical modeling) are of particular concern. For example, in order to interpret and generalize conclusions to our population of Medicare patients, it may be necessary for studies to match or stratify their intervention and control groups by patient age or co-morbidities.

Methodological strength is, therefore, a multidimensional concept that relates to the design, implementation and analysis of a clinical study. In addition, thorough documentation of the conduct of the research, particularly study selection criteria, rate of attrition and process for data collection, is essential for CMS to adequately assess and consider the evidence.

Generalizability of Clinical Evidence to the Medicare Population

The applicability of the results of a study to other populations, settings, treatment regimens and outcomes assessed is known as external validity. Even well designed and well-conducted trials may not supply the evidence needed if the results of a study are not applicable to the Medicare population. Evidence that provides accurate information about a population or setting not well represented in the Medicare program would be considered but would suffer from limited generalizability.

The extent to which the results of a trial are applicable to other circumstances is often a matter of judgment that depends on specific study characteristics, primarily the patient population studied (age, sex, severity of disease and presence of co-morbidities) and the care setting (primary to tertiary level of care, as well as the experience and specialization of the care provider). Additional relevant variables are treatment regimens (dosage, timing and route of administration), co-interventions or concomitant therapies, and type of outcome and length of follow-up.

The level of care and the experience of the providers in the study are other crucial elements in assessing a study's external validity. Trial participants in an academic medical center may receive more or different attention than is typically available in non-tertiary settings. For example, an investigator's lengthy and detailed explanations of the potential benefits of the intervention and/or the use of new equipment provided to the academic center by the study sponsor may raise doubts about the applicability of study findings to community practice.

Given the evidence available in the research literature, some degree of generalization about an intervention's potential benefits and harms is invariably required in making coverage determinations for the Medicare population. Conditions that assist us in making reasonable generalizations are biologic plausibility, similarities between the populations studied and Medicare patients (age, sex, ethnicity and clinical presentation) and similarities of the intervention studied to those that would be routinely available in community practice.

A study's selected outcomes are an important consideration in generalizing available clinical evidence to Medicare coverage determinations. One of the goals of our determination process is to assess net health outcomes. These outcomes include resultant risks and benefits such as increased or decreased morbidity and mortality. In order to make this determination, it is often necessary to evaluate whether the strength of the evidence is adequate to draw conclusions about the direction and magnitude of each individual outcome relevant to the intervention under study. In addition, it is important that an intervention's benefits are clinically significant and durable, rather than marginal or short-lived.

If key health outcomes have not been studied or the direction of clinical effect is inconclusive, we may also evaluate the strength and adequacy of indirect evidence linking intermediate or surrogate outcomes to our outcomes of interest.

Assessing the Relative Magnitude of Risks and Benefits

In general, an intervention is not reasonable and necessary if its risks outweigh its benefits. Among other things, CMS considers whether reported benefits translate into improved net health outcomes. CMS places greater emphasis on health outcomes actually experienced by patients, such as quality of life, functional status, duration of disability, morbidity and mortality, and less emphasis on outcomes that patients do not directly experience, such as intermediate outcomes, surrogate outcomes, and laboratory or radiographic responses. The direction, magnitude, and consistency of the risks and benefits across studies are also important considerations. Based on the analysis of the strength of the evidence, CMS assesses the relative magnitude of an intervention or technology's benefits and risk of harm to Medicare beneficiaries.

Appendix C
MCAC Questions and Scoring Summary

The panel voted on the following questions. The possible scores range from 1 (low) to 5 (high).

FOR CARDIORESPIRATORY MEASURES ONLY:

Question 1. How well does the evidence address the effectiveness of this type of unattended portable multichannel home sleep testing devices as an alternative to facility based polysomnography in the diagnosis of obstructive sleep apnea, or OSA?

Score Average: 2.375

Question 2. How confident are you in the validity of the scientific data for the following outcomes?

a. Acquisition of interpretable data? Score Average: 2.5
b. Ability to accurately diagnose OSA (sensitivity)? Score Average: 2.8125
c. Ability to accurately identify those without OSA (specificity)? Score Average: 2.8125

Question 3. How likely is it that these home sleep testing devices will be as good as or better than facility-based polysomnography for the following outcomes?

a. Acquisition of interpretable data? Score Average: 2.375
b. Ability to accurately diagnose OSA (sensitivity)? Score Average: 2.75
c. Ability to accurately identify those without OSA (specificity)? Score Average: 2.75

Question 4a. How confident are you that these sleep testing devices are as accurate in the diagnosis of obstructive sleep apnea as is a facility based test? Score Average: 2.5

Question 4b. How confident are you that use of these sleep testing devices in the diagnosis of obstructive sleep apnea will lead to similar or improved health outcomes measured either directly or indirectly through changes in patient management as compared to a facility based test? Score Average: 3.125

Question 4c. How confident are you that these sleep testing devices are as accessible as is a facility based test for the diagnosis of obstructive sleep apnea? Score Average: 4.25

Question 5. Based on the literature presented, how likely is it that the evidence addressing the diagnosis of OSA utilizing these sleep testing devices can be generalized to:

a. The Medicare population (aged 65+)? Score Average: 2.5
b. Providers (facilities/physicians) in community practice? Score Average: 2.57

SLEEP AND RESPIRATORY PARAMETERS

Question 1. How well does the evidence address the effectiveness of this type of unattended portable multichannel home sleep testing devices as an alternative to facility based polysomnography in the diagnosis of obstructive sleep apnea, or OSA? Score Average: 2

Question 2. How confident are you in the validity of the scientific data for the following outcomes:

a. Acquisition of interpretable data? Score Average: 2.375
b. Ability to accurately diagnose OSA (sensitivity)? Score Average: 2.25
c. Ability to accurately identify those without OSA (specificity)? Score Average: 2.625

Question 3. How likely is it that these home sleep testing devices will be as good as or better than facility-based polysomnography for the following outcomes?

a. Acquisition of interpretable data? Score Average: 2
b. Ability to accurately diagnose OSA (sensitivity)? Score Average: 2.75
c. Ability to accurately identify those without OSA (specificity)? Score Average: 3

Question 4a. How confident are you that these sleep testing devices are as accurate in the diagnosis of obstructive sleep apnea as is a facility based test? Score Average: 2.5

Question 4b. How confident are you that use of these sleep testing devices in the diagnosis of obstructive sleep apnea will lead to similar or improved health outcomes measured either directly or indirectly through changes in patient management as compared to a facility based test? Score Average: 2.875

Question 4c. How confident are you that these sleep testing devices are as accessible as is a facility based test for the diagnosis of obstructive sleep apnea? Score Average: 4.25

Question 5. Based on the literature presented, how likely is it that the evidence addressing the diagnosis of OSA utilizing these sleep testing devices can be generalized to:

a. The Medicare population (aged 65+)? Score Average: 2.375
b. Providers (facilities/physicians) in community practice? Score Average: 2.43

---

[1] Young, et al. (1993)

[2] Peppard, et al. (1997); Lavie, et al. (2000); Nieto, et al. (2000); Weiss, M.D. Cardiovascular and cerebrovascular effects of sleep apnea. UpToDate (2004)

[3] Strollo, et al. (1996); Bradley, et al. (1985)

[4] Lombard, et al. (1985); Loube, et al. (1999)

[5] Gould, et al. (1988)

[6] Office of Health Technology Assessment (1986)

[7] Jennet, et al. (1984)

[8] 8 Hulley, et al. Designing Clinical Research. 2001.

[9] The TA report is available at: http://www.cms.hhs.gov/mcd/viewtechassess.asp?id=110. A complete description of search strategies and article inclusion/exclusion criteria is included.

[10] Effectiveness of Portable Monitoring Devices for Diagnosing Obstructive Sleep Apnea: Update of a Systematic Review. Submitted to AHRQ by RTI International- UNC Evidence Based Practice Center.

[11] Flemons, et al. Home diagnosis of sleep apnea: A systematic review of the literature. An evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society. Chest 2003;124:1543-79.

[12] Dingli, et al. (2003), Reichert, et al. (2003)

[13] Portier, et al. (2000); Orr, et al. (1994); Mykytyn, et al. (1999); Iber, et al. (2004); Fry, et al. (1998)

[14] Ancoli-Israel, et al. (1997); Para, et al. (1997); Whittle, et al. (1997)

[15] Dingli, et al. (2003); Reichert, et al. (2003); White, et al. (1995); Redline, et al. (1991)

[16] Marrone, et al. (2001); Ficker, et al. (2001); Ballester, et al. (2000); Claman, et al. (2001); Verse, et al. (1998); Esnaola, et al., (1996); Calleja, et al. (2002); Zucconi, et al. (1996); Man, et al. (1995)

[17] Flemons, et al. Measuring Agreement Between Diagnostic Devices. Chest 2003; 124:15 35-42

[18] Whittle (1997), Parra (1997), Dingli (2003)

[19] Reichert (2003), White (1995), and Ancoli-Israel (1997)

[20] Dingli (2003) and Redline (1991)

[21] Millman R and Kramer N. Polysomnography in the diagnostic evaluation of sleep apnea. UpToDate 2004.

[22] Tsai, et al. (2001)

Back to Top

Bibliography

1. Redline S and Young T. Epidemiology and natural history of obstructive sleep apnea. Ear Nose Throat J 1993;72(1):20-1, 24-6. Review.

2. Young T, Peppard P, Palta M, et al. Population-based study of sleep-disordered breathing as a risk for hypertension. Arch Intern Med. 1997;157(15):1746-52.

3. Lavie P, Herer P, and Hoffstein V. Obstructive sleep apnoea syndrome as a risk factor for hypertension: population study. BMJ. 2000;320(7233):479-82.

4. Gottlieb DJ, Whitney CW, Bonekat WH, et al. Relation of sleepiness to respiratory disturbance index: the Sleep Heart Health Study. Am J respire Crit Cared Med. 1999;159(2):502-7.

5. Strollo PJ Jr and Rogers RM. Obstructive sleep apnea. N Engl J Med. 1996;334(2):99-104.

6. Bradley TD and Phillipson EA. Pathogenesis and pathophysiology of the obstructive sleep apnea syndrome. Med Clin North Am. 1985;69(6):1169-85.

7. Lombard RM Jr and Zwillich CW. Medical therapy of obstructive sleep apnea. Med Clinic North Am 1985o;69(6)1317-35.

8. Hulley, et al. Designing Clinical Research. 2001

9. The Technology Assessment submitted to AHRQ by RTI September 1, 2004. WWW.CMS.HHS.GOV/Coverage.

10. Effectiveness of Portable Monitoring Devices for Diagnosing Obstructive Sleep Apnea: Update of a Systematic Review. Submitted to AHRQ by RTI International- UNC Evidence Based Practice Center.

11. Flemons, et al. Home diagnosis of sleep apnea: A systematic review of the literature. An evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society. Chest 2003; 124:1543-79.

12. Dingli K, Coleman EL, Vennelle M, et al. Evaluation of a portable device for diagnosing the sleep apnoea/hypopnoea syndrome. Eur Respir J. 2003; 21:253-9.

13. Reichert J, Bloch D, Cundiff E, Votteri B. Comparison of the NovaSom QSG™, a new sleep apnea home-diagnostic system, and polysomnography. Sleep Medicine 4 2003; 213-8.

14. Portier F, Portmann A, Czernichow P, et al. Evaluation of Home versus Laboratory Polysomnography in the Diagnosis of Sleep Apnea Syndrome. Am J Respir Crit Care Med 2000; 162:814-18.

15. Orr W, Eiken T, Vernon P, et al. A Laboratory Validation Study of a Portable System for Remote Recording of Sleep-related Respiratory Disorders. Chest 1994; 105:160-62.

16. Mykytyn I, Sajkov D, Neill A, et al. Portable Computerized Polysomnography in the Attended and Unattended Settings. Chest 1999;115:114-22.

17. Iber C, Redline S, Gilpin A, et al. Polysomnography Performed in the Unattended Home Versus the Attended Laboratory Setting- Sleep Heart Health Study Methodology. Sleep 2004; 27(3):536-40.

18. Fry J, DiPhillipo M, Curran K, et al. Full Polysomnography in the Home. Sleep 1998; 21:635-42.

19. Ancoli-Israel S, Mason W, Coy TV, et al. Evaluation of sleep disordered breathing with unattended recording: the Nightwatch System. Journal of Medical Engineering and Technology 1997; 21(1):10-14

20. Parra O, Garcia-Esclasans N, Montserrat JM, et al. Should patients with sleep apnoea/hypopnoea syndrome be diagnosed and managed on the basis of home sleep studies? Eur Respir J 1997; 10:1720-24.

21. Whittle AT, Finch SP, Mortimore IL, et al. Use of home sleep studies for diagnosis of the sleep apnoea/hypopnoea syndrome. Thorax 1997; 52:1068-73.

22. White D, Gibb T, Wall J, Westbrook P. Assessment of Accuracy and Analysis Time of a Novel Device to Monitor Sleep and Breathing in the Home. Sleep 1995; 18(2):115-26.

23. Redline S, Tosteson T, Boucher MA, and Millman R. Measurement of Sleep-related Breathing Disturbances in the Epidemiologic Studies. Chest 1991; 100:1281-86.

24. Marrone O, Salvaggio A, Insalaco G, et al. Evaluation of the POLYMESAM® system in the diagnosis of obstructive sleep apnea syndrome. Monaldi Arch Chest Dis 2001; 56(6):486-90

25. Ficker JH, Wiest GH, Wilpert J, et al. Evaluation of a Portable Recording Device for Use in Patients with Suspected Obstructive Sleep Apnoea. Respiration 2001;68:307-12.

26. Ballester E, Solans M, Villa X, et al. Evaluation of a portable respiratory device for detection apnoeas and hypopnoeas in subjects from a general population. Eur Respir J 2000; 16: 123-27.

27. Claman D, Murr A, and Trotter K. Clinical validation of the Bedbugg™ in detection of obstructive sleep apnea. Otolaryngol Head Neck Surg 2001; 125:227-30.

28. Verse T, Wolfgang P, Junge-Hülsing B, et al. Validation of the POLY-MESAM Seven-Channel Ambulatory Recording Unit. Chest 2000;117:1613-18.

29. Esnaola S, Duran J, Infante-Rivard C, et al. Diagnostic accuracy of a portable recording device (MESAM IV) in suspected obstructive sleep apnoea. Eur Respir J 1996; 9: 2597-2605.

30. Calleja JM, Esnaola S, Rubio R, et al. Comparison of a cardiorespiratory device versus polysomnography for diagnosis of sleep apnoea. Eur Respir J 2002; 20:1505-10.

31. Zucconi M, Fernini-Strambi L, Castronovo V, et al. An unattended device for sleep-related breathing disorders: validation study in suspected obstructive sleep apnoea syndrome. Eur Respir J 1996; 9:1251-56.

32. Man G, Kang B, et al. Validation of a Portable Sleep Apnea Monitoring Device. Chest 1995; 108:388-93.

33. Flemons WW, and Littner MR. Measuring Agreement Between Diagnostic Devices. Chest 2003; 124:1535-42.

34. Boyer S and Kapur V. Role of portable sleep studies for diagnosis of obstructive sleep apnea. Curr Opin Pulm Med 2003; 9:465-70.

35. Chesson AL Jr, Berry RB, and Pack A. Practice parameters for the use of portable monitoring devices in the investigation of suspected obstructive sleep apnea in adults. Sleep 2003; 26:907-13.

36. Loube DI. Technologic advances in the treatment of obstructive sleep apnea syndrome. Chest 1999; 116(5):1426-33

37. Gould GA, Whyte KF, Rhind GB, et al. The sleep hypopnea syndrome. Am Rev Respir Dis. 1998;137(4):895-8.

Back to Top