Effectiveness of Portable Monitoring Devices for Diagnosing Obstructive Sleep Apnea: Update of a Systematic Review

Final Report

Submitted to:
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, Maryland 20852

Submitted by:
RTI International
3040 Cornwallis Road
P.O. Box 12194
Research Triangle Park, North Carolina 27709

Contract No. 290-02-0016

RTI Project No. 8452.001.005

September 1, 2004

Technology Assessment

Technology Assessment Program

Effectiveness of Portable Monitoring Devices for Diagnosing Obstructive Sleep Apnea: Update of a Systematic Review

September 1, 2004

Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, Maryland 20850

Effectiveness of Portable Monitoring Devices for Diagnosing Obstructive Sleep Apnea: Update of a Systematic Review

Final Report

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-02-0016

Prepared by:
RTI-UNC Evidence-based Practice Center
3040 Cornwallis Road
P.O. Box 12194
Research Triangle Park, North Carolina 27709

Investigators
Linda Lux, MPA
Brian Boehlecke, MD, MSPH
Kathleen N. Lohr, PhD

Structured Abstract

Context: Obstructive sleep apnea (OSA) is a serious public health problem. Approximately 2 percent to 4 percent of middle-aged women and men, respectively, have this condition; the majority are undiagnosed. Undiagnosed and thus untreated, OSA is associated with significant morbidity and mortality. Effective treatment modalities should not be applied without an accurate diagnosis of OSA, but medical history and physical examination are insufficient to establish the diagnosis or its severity. Using the accepted reference standard test – attended, in-laboratory polysomnography (PSG) – can be expensive and involve long waiting times for studies, so various groups have developed portable technologies to classify patients in terms of the presence or absence of OSA and, for the former, level of severity. Such devices are intended for use in sleep laboratories or in the home.

Objectives: We updated a 2002-2003 systematic review of OSA diagnostic testing to address the key questions of how portable sleep testing devices compared to PSG in diagnosing OSA and, assuming equivalent effectiveness, what sleep and physiologic factors and what patient and technician conditions were important to measure or have in

place. The Centers for Medicare and Medicaid Services commissioned the Agency for Healthcare Research and Quality to provide a technology assessment that addressed the following:

1. 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, 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?

Data Sources: We searched for studies published since the original review (i.e., from 2002 on) in MEDLINE, The Cochrane Library, the National Guidelines Clearinghouse, and the International Network of

Agencies for Health Technologies Assessment (INAHTA) database; we also handsearched bibliographies of included articles. In MEDLINE, we used the following main terms in various combinations: polysomnography, oximetry, physiologic monitoring, and sleep apnea (with limits of human, adults, and English language); we refined searches using the terms airway resistance, upper airway resistance syndrome, respiratory disturbance index, autoset, snoring, and respiratory events related arousals as well as reproducibility of results, predictive value of tests, and sensitivity and specificity.

Study Selection: We included studies of humans, both sexes, ages 18 and over, with any diagnosis of OSA; studies of any type of portable device used for diagnosis that also included a reference standard test (PSG or another acceptable test for diagnosing OSA); studies in which each analysis group, after the end of the study, included at least 10 subjects; and studies published in English. Specifically excluded were studies in which results from portable devices were not compared with results from PSG. Also excluded were reviews, meta-analyses, case reports, abstracts, letters, and editorials.

Data Extraction: One investigator recorded abstracted data onto data abstraction forms used for the original review and created detailed

evidence tables. A second investigator checked entries against the original articles. One investigator assigned initial classifications for level of evidence and presence or absence of eight quality indicators and a second investigator reviewed these; disagreements were resolved by consensus discussion. A third investigator combined level of evidence and quality indicators into a summary quality grade; the other investigators reviewed these grades, with differences resolved by consensus.

Data Synthesis: We identified 172 unique titles and abstracts from the literature searches, and excluded 157 articles as not meeting inclusion criteria; reasons included the fact that PSG studies were not performed on all patients, that the portable device was an electroencephalogram (EEG), and that the study assessed a telemedicine approach that did not compare a portable device to the PSG. We obtained 15 articles for full review and retained 12 for inclusion here.

These 12 studies fell into four categories: Type 3 devices used in laboratory settings (four studies); Type 3 devices tested in homes whether or not they were also tested in facilities (two studies); Type 4 devices in laboratory settings (six studies); and Type 4 devices tested in homes (whether or not in facilities, three studies). Type 3 devices include a minimum of four channels and must monitor at least two channels of

respiratory movement or respiratory movement and airflow, and heart rate or ECG and oxygen saturation to define an event; generally, no electroencephalogram (EEG) signals are monitored. Level 4 devices include only one or two channels of physiologic signals and generally use only one channel (either saturated oxygen or airflow) to define a sleep-disordered breathing event; no EEG signals are monitored.

Most articles provided only comparisons of the results from portable monitoring done simultaneously with full PSG in the laboratory, i.e., "a side-by-side" study. The in-laboratory simultaneous studies, which used technologies identical or similar to those in the previous review, produced sensitivity and specificity results for diagnosing OSA similar to those reported earlier; that is, the newer studies produced no meaningful changes in the level or quality of evidence for the effectiveness for home monitoring devices in diagnosing OSA. Only four of these studies (two of Type 3 and two of Type 4 devices) were graded good or fair quality.

Ultimately, we focused on the five studies with in-home testing, because the questions we were asked concerned the effectiveness of unattended monitoring in the home. Four in-home studies employed technologies similar or identical to those reviewed before; of these, two studies (one of good quality, one poor) used Type 3 devices and two (one

of fair quality , one poor) used Type 4 devices. Reported sensitivity and specificity values were similar to those from older studies, so the newer studies yielded no major information that would change the previous basic conclusions about portable devices used in the home. The one in-home study using a new technology, of fair quality, produced likelihood ratios that indicated that the test had little effect in changing pretest probabilities of the presence or absence of OSA. Reported data loss in the home studies ranged from a low of 3 percent to a high of 33 percent, in a subgroup of patients who did their own hookup. Automated scoring appeared to agree less closely with the reference standard than manual scoring. Internal validity of the five in-home studies was mixed: one study of good quality, two of fair quality, and two of poor quality. In terms of external validity, the patient populations were mostly male, middle-aged, and with high pretest probabilities of OSA; comorbidities were generally not specified or taken into account in analyses. Finally, these studies typically did not evaluate the accuracy of clinical management decisions based on portable results compared to those based on the reference standard.

Conclusions: This newer body of evidence does not materially change earlier findings regarding in-home devices for diagnosing OSA. Choices of cutoffs for determining OSA by AHI or RDI differed widely

across these studies, making cross-study comparisons impossible. The better studies yielded sensitivity and specificity values (or LRs) that provided modes changes in the probability of OSA over the pretest probability. In studies that directly compared automated versus manual scoring from home monitoring devices, manual scoring correlated better with data from laboratory PSG than did automated scoring.

Table of Contents

Structured Abstract i
Chapter 1. Introduction 1
Background 1
Evaluating the Role of Home Testing of Obstructive Sleep Apnea 3
The Original 2002 RTI-UNC Systematic Evidence Review 4
Purpose of this Updated Evidence Report 8
Chapter 2. Methods 11
Literature Search Strategy 11
Data Abstraction 13
Quality Ratings 14
Chapter 3. Results 18
General Overview 19
Types of Devices 19
Patient Populations 20
Location of Studies 20
Implementation of Testing Studies 21
Scoring and Interpreting Portable Monitor Results 22
Sponsorship of Studies 23
Quality of Studies and Strength of Evidence 23
Important Findings from Reviewed Studies 24
Portable Devices Studied in the Laboratory 25
Portable Devices Studied in the Home 27
Chapter 4. Discussion 39
Introduction 39
Critical Issues in Evaluating Portable Monitor Studies 40
Portable Testing in Laboratories or in Homes 41
Data Loss 41
Manual or Automated Scoring 42
Core Aspects of Clinical Management 45
Other Clinical Issues 46
Co-existing Conditions 48
Similarity of Update and Prior Findings 49
Issues of Internal Validity of Reviewed Studies 50
External Validity or Generalizability 54
Overall Generalizability 54
Generalizability to the Medicare Beneficiary Population 55
Summary 57

References 60

Appendices

Appendix A. Acknowledgments A-1
Appendix B. Data Abstraction Form B-1
Appendix C. Evidence Tables C-1

List of Tables

Table 1. Inclusion and Exclusion Criteria 11
Table 2. Quality Grades Based on Levels of Evidence and Scores on Quality Indicators 17
Table 3. Twelve Included Studies with Type of Device, Site of Test, and Quality Grade 34
Table 4. Summary of Studies of Portable Monitoring Devices Conducted in the Home 35

Chapter 1. Introduction

Background

Obstructive sleep apnea (OSA, sometimes characterized as obstructive sleep apnea syndrome) is now recognized to be a significant and serious public health problem. Researchers have estimated that approximately 2 percent to 4 percent of middle-aged women and men, respectively, have this condition; the majority (approximately 80 percent to 90 percent in one study[1]) remain undiagnosed. Undiagnosed and thus untreated, OSA is associated with significant morbidity and mortality, including excessive daytime somnolence, increased risk of automobile crashes, hypertension, cardiovascular disease, stroke, and metabolic abnormalities.

Effective treatment modalities are available, primarily nocturnal continuous positive airway pressure (CPAP) and in some instances surgical procedures or dental appliances. These treatments, however, are expensive and have potential side effects, so they should not be applied without an accurately established diagnosis of OSA.

Medical history and physical examination can provide an estimate of the likelihood of OSA, but they are not sufficient to establish the diagnosis

or its severity. Therefore, patients suspected of having this condition must be evaluated with a diagnostic test that can provide a significant increase ("rule in") or decrease ("rule out") in the likelihood of the condition so that proper management can be implemented.

Using the reference standard polysomnography (PSG), which is a facility-based diagnostic intervention, is expensive. Because of limited facilities, waiting time for studies after the diagnosis is suspected on clinical grounds has been excessive in many areas of the country. Thus, various groups have attempted to develop portable technologies that can accurately classify patients as either having a very low likelihood of OSA and thus not need a PSG or having a very high likelihood of OSA and for whom management with a CPAP titration or other procedure should be initiated. The goal is to reduce the need for expensive laboratory testing while increasing the rapidity of diagnosis and initiation of appropriate management.

The first step in determining whether a portable monitoring device can achieve this goal is to determine its accuracy in characterizing the presence and severity of sleep-disordered breathing events relative to the reference standard PSG in a controlled study. This is the focus of most of the research papers published on portable monitors for OSA, and it was the

main factor considered in the last systematic evidence review (described below). However, other considerations are also important in the overall assessment of whether the current technologies will be cost effective and provide adequate accuracy of diagnosis if applied to a large population of patients in unattended settings. These issues are examined further in the Discussion chapter of this report.

Evaluating the Role of Home Testing of Obstructive Sleep Apnea

As noted, various portable devices have been developed over the past decade that are meant to be used as screening tools or replacements for the labor-intensive, complex, expensive, laboratory- or facility-based PSG for the evaluation of patients suspected of having OSA.

The American Sleep Disorders Association classified monitors used in diagnostic testing for sleep apnea into four types.[2] Attended PSG, Type 1, is the gold standard, and the portable monitors fall into three types (2, 3, and 4) with fewer physiologic signals monitored in each subsequent type. The levels are briefly defined below to clarify the differences between them:

- Type 1: Measures, at a minimum, eight channels ─ electroencephalogram (EEG), electro-oculogram (EOG), electrocardiogram (ECG), chin electromyogram (EMG), airflow,

respiratory "effort," oxygen saturation (SaO_2), and body position; it is attended in a laboratory setting.

- Type 2: Monitors a minimum of seven channels including EEG, EOG, chin EMG, ECG or heart rate, airflow, respiratory effort, and SaO_2. This allows for sleep staging and measurement of total sleep time, and this information can be used to determine in the number of sleep-disordered breathing events per hour of sleep (e.g., the apnea/hypopnea index).
- Type 3: Includes a minimum of four channels and must monitor at least two channels of respiratory movement or respiratory movement and airflow to define an event; generally, no EEG signals are monitored.
- Level 4: Includes only one or two channels of physiologic signals and generally uses only one channel (eitherSaO_2 or airflow) to define a sleep-disordered breathing event; no EEG signals are monitored.

The Original 2002 RTI-UNC Systematic Evidence Review

In 2002, RTI International (RTI) and the University of North Carolina at Chapel Hill (UNC) completed a systematic review of published articles on home diagnostic testing for sleep apnea in collaboration with three

professional organizations: the American Academy of Sleep Medicine (ASSM), the American College of Chest Physicians (ACCP), and the American Thoracic Society (ATS). The RTI-UNC team conducted the literature search and prepared the evidence tables. Members of the three organizations analyzed the data and prepared three key publications.

The full evidence review was subsequently published in the journal *Chest* in October 2003.[3] An article in the same issue of *Chest* measured agreement between diagnostic devices.[4] Finally, new recommended clinical practice guidelines (practice parameters) from these three professional organizations for the use of portable monitoring devices in the investigation of suspected obstructive sleep apnea (SOSA) in adults was published in *Sleep.*[5] An executive summary of the systematic review and practice parameters listed above was published in 2004 in the *American Journal of Respiratory and Critical Care Medicine.*[6]

The 2002 review compared all of the portable devices to the Type 1, attended, in-laboratory PSG. Although the accuracy of a single-night PSG in determining the presence or absence of clinically significant OSA does have certain limitations (see Discussion section of this report), to date no better standard has emerged. Therefore, this review also compares portable home monitoring devices to the PSG.

The most widely used measure to define the presence and severity of OSA by PSG is the apnea/hyponea index (AHI). Apneas in this calculation are events with complete cessation of airflow; hypopneas are events with decreases in airflow without complete cessation but with associated decreases in SaO_2 or EEG arousals (or both) depending on the definition used by the researcher or clinician. The AHI is the number of disordered breathing events per hour of sleep calculated from the total number of apneas and hypopneas. Most studies use a lower cutoff level for the AHI to define the presence of OSA by PSG; increasing levels of AHI indicate increasing severity of OSA.

If portable monitoring does not allow for determination of sleep time, then the AHI cannot be calculated. Instead, researchers calculate the number of disordered breathing events per hour in bed or per hour of monitoring time and report this as the respiratory disturbance index (RDI). The RDI can be compared with the same measure calculated from a PSG.

In our original (2002) evidence report, we included 51 studies. Of these, four studies were of Type 2 devices, 12 of Type 3, and 35 of the Type 4. The joint ATS/ACCP/AASM summary of the evidence review and practice parameters stated that data were not adequate to recommend the clinical use of Type 2 portable monitors in either attended or unattended

settings. Neither sensitivity nor specificity data were available; moreover, the number of studies of Type 2 devices was small. Overall, the level of evidence was low. Some Type 3 monitors appeared to be potentially acceptable in the attended laboratory setting, but six limitations were noted:

1. Careful review of raw data is necessary (e.g., manual or a combination of automatic and manual scoring).
2. The devices should be used only in populations that have been studied (e.g., those in a sleep clinic population) and should not be applied as generalized screening or in populations with significant comorbidity such as chronic obstructive pulmonary disease or congestive heart failure.
3. AHI in these devices tends to underestimate the PSG-defined AHI because these devices do not measure sleep time.
4. Symptomatic patients with a nondiagnostic or negative Type 3 study should undergo definitive evaluation to determine the cause of symptoms; if a sleep disorder requiring a sleep study is still a clinical consideration, then a full attended PSG should be used.
5. Patients with a positive Type 3 study need a subsequent PSG if CPAP titration is needed.

6. Type 3 portable monitors are not recommended for split-night studies.[6]

In essence, these Type 3 devices were not recommended for unattended use in the home. Type 4 devices were not recommended for diagnostic use or to assess the probability that a patient may or may not have OSA.

Overall, portable devices were not recommended either for general screening or for patients with certain comorbid conditions. Manual scoring by "physicians with specific sleep training and familiarity with the devices and their limitations" was recommended rather than use of the automated scoring available for some portable devices.

Purpose of this Updated Evidence Report

RTI is now assisting the Agency for Healthcare Quality and Research (AHRQ) to develop a summary of the available clinical and scientific evidence on home diagnosis of OSA since the last review in 2002. The Centers for Medicare and Medicaid Services (CMS) commissioned AHRQ to provide a technology assessment in preparation for a Medicare Coverage Advisory Committee (MCAC) meeting on September 28, 2004, at which the MCAC will review the evidence.

CMS has provided one key question for this evidence review, with two subquestions based on the results of the main question:

1. 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, 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?

In this updated review we report on 12 studies published since the last full review in 2002 that present data on the use of portable monitoring devices for evaluation of OSA.

Chapter 2 describes our methods for this update. Chapter 3 presents our findings from the updated evidence review and Chapter 4 discusses the implications of our results. Acknowledgments can be found in Appendix A; our data abstraction form is in Appendix B; evidence tables (one per study) appear in Appendix C.

Chapter 2. Methods

Literature Search Strategy

We updated the literature search used for the previous systematic review that we had completed in 2002, using the same inclusion and exclusion criteria. Both the Agency for Healthcare Research and Quality (AHRQ) and the Centers for Medicare and Medicaid Services (CMS) reviewed and approved these criteria (Table 1).

Table 1. Inclusion and Exclusion Criteria

Inclusion Criteria	
Patient populations	Human, both sexes, ages 18 and over, with ANY diagnosis of obstructive sleep apnea (OSA)
Types of OSA diagnostic device	Portable device used for diagnosis AND polysomnography or other acceptable test used for diagnosis of OSA
Sample size	After completion of study, each analysis group is greater than or equal to 10 subjects
Language	Study published in English
Exclusion Criteria	
Patient populations	Children (birth through 17 years)
Study types	Reviews, meta-analyses, case reports, abstracts letters, editorials
Language	Other than English

Our search concentrated on articles published since January 1, 2002, a date that coincides with the end of the coverage period for the most recent evidence review.

The basic Medical Subject Headings (MeSH) used in our updated search were polysomnography (PSG), oximetry, physiologic monitoring, and sleep apnea (with limits, as noted in Table 1, of human, adults, and English language). We also included the following terms to help us refine our search: airway resistance, upper airway resistance syndrome, respiratory disturbance index, autoset, snoring, and respiratory events-related arousals. Finally, to understand the scope of the impact of screening on the societal burden of sleep apnea, we added several other search terms: reproducibility of results, predictive value of tests, and sensitivity and specificity.

Our search yielded 172 items with titles and abstracts. Of these, we determined that 15 were eligible for full retrieval and review. We were able to obtain 13 of these articles, but after review, excluded three. We also conducted a handsearch and identified two additional studies, for a total of 12 studies for inclusion in the evidence tables. Some articles identified through handsearching did not meet the criteria for inclusion in the evidence tables, but they did contain information relevant to the key questions and will be referenced in the discussion chapter.

A search of the International Network of Agencies for Health Technologies Assessment (INAHTA) database for Diagnostic Procedures

and Screening yielded no relevant reports to review. The National Guidelines Clearinghouse has two guidelines: "Practice parameters for the use of portable monitoring devices in the investigation of suspected sleep apnea in adults"[5] by the three professional societies that published manuscripts on our 2002 review, discussed in Chapter 1; and "Diagnosis and treatment of obstructive sleep apnea" by the Institute for Clinical Systems Improvement, also published in 2003.[7] We reviewed both these documents for potential studies to include in our review and used them as background material for this review. A search of the Cochrane Collaboration Database turned up a number of hits, but further investigation indicated that the studies or reviews so identified were either not relevant to this review or outdated.

Data Abstraction

We used a data abstraction template from the earlier review to retrieve and enter the relevant information from the articles that presented new evidence (Appendix B). The form includes a general description of study design and outcomes for each arm (device or control/comparison group) studies, including sample size; patient characteristics, types, frequency, and duration of monitoring; outcomes measures; and the statistical significance of differences.

We created evidence tables organized by the level of device (see Chapter 1). The first part of the evidence table describes the diagnostic device and how it is applied/used and the nature of the target population (demographic characteristics, presence of conditions related to sleep apnea or symptoms of sleep apnea). The second part describes the research design and conduct; the third part describes the study outcomes. To ensure consistency, we developed standardized abbreviations and conventions for describing particular pieces of information. The evidence tables can be found in Appendix C.

The evidence tables are formatted to capture detailed, relevant characteristics of each study. They are also flexible enough to accommodate study designs ranging from randomized controlled trials (RCTs) to quasi-experimental and observational studies.

Quality Ratings

Rating the quality of individual studies is a critical element of any systematic review. For this step, we applied the same quality rating scheme as we used in the 2002 review for grading by the AASM/ACCP/ATS Evidence Review Committee,[3] which was based on levels of evidence and quality indicator scores.

The levels of evidence were followed the approach published by Sackett et al.[8] and the four levels were defined as follows:

I. Blinded comparison, consecutive patients, reference standard (in this case PSG) performed on all patients;

II. Blinded comparison, *nonconsecutive* patients, reference standard performed on all patients;

III. Blinded comparison, consecutive patients, reference standard *not performed* on all patients; and

IV. Reference standard *not applied blindly* or independently.

For this updated review, we included no Level III studies because of the inclusion criterion that specified that the portable device had to be compared to the PSG, the best-known reference standard at this time. We thus reviewed each study for inclusion of the following three items to determine the level of evidence:

1. Subjects selected consecutively or randomly;
2. Reference standard done on all subjects;
3. Both tests scored blindly.

If the study did not report on any of the items that make up the assigned level of evidence, we assigned a value of “no.”

For the quality indicators score, we used the original eight indicators developed during the 2002 review and applied them to each study in this review:

- Prospective study;
- Portable device tested outside the sleep laboratory;
- Random order of allocation of subjects to the PSG or portable test first;
- Low data loss, specifically ≤ 10 percent of flow/SaO_2/EEG, as applicable;
- High percentage of completions (≥90% of those entered);
- PSG methodology fully described;
- Portable methodology fully described;
- Portable scoring fully described.

As with the levels of evidence, if the study did not report on an indicator, then we assigned a "no" value; for the third item above (random order), we scored "not applicable" when the portable and the PSG tests were done simultaneously. Using the results of the number of quality indicators that were met for each study, we calculated an overall quality indicator score based on the categories defined below:

Quality rating:

A. Zero or only one quality indicator NOT met;

B. Two quality indicators NOT met;

C. Three quality indicators NOT met; or

D. Four or more quality indicators NOT met.

For ease of interpretation, we developed a summary quality grade for each article. We used the above methodology and then arbitrarily assigned a grade for each combination of level of evidence and quality score. Table 2 below shows the specific mapping of each category for the levels of evidence and quality indicator scores to each individual quality grade.

Table 2. Quality Grades Based on Levels of Evidence and Scores on Quality Indicators

Grades	Levels of Evidence and Quality Indicator Scores			
	Level I	Level II	Level III	Level IV
Good	A, B		NA	
Fair	C	A, B,C	NA	A, B
Poor	D	D	NA	C, D

NOTE: Scores on eight quality indicators: A, 0 –1 missed; B, 2 missed; C, 3 missed; D, 4 or more missed. Level III evidence was not applicable for this updated review. No Level II or IV studies could be considered good evidence (see text for discussion).

Chapter 3. Results

The updated literature search described in Chapter 2 initially produced 172 titles and abstracts that were reviewed against the inclusion and exclusion criteria. We obtained 15 articles for full review and ultimately retained 12 articles as relevant to the key questions detailed in Chapter 1. They date from 2001 through 2004.[9-20]

We excluded studies for a variety of reasons. These typically included the fact that polysomnography (PSG) studies were not performed on all patients, that the portable device was only an electroencephalogram (EEG), and that the study assessed a telemedicine approach that did not compare a portable device to the PSG.

The remainder of this chapter describes the studies in terms of type of device, setting, populations, and similar design characteristics and summarizes some qualitative findings on this evidence base. We then present specific information drawn from five articles in which the portable device was tested in a home setting. Detailed information on all 12 studies (Table 3) can be found in the evidence tables in Appendix C.

As noted in Chapter 1, the chief issues for this review concern the diagnostic test properties of unattended, portable sleep testing devices

compared with those of facility-based PSG in diagnosing obstructive sleep apnea (OSA). Two subsidiary questions, predicated on similar effectiveness, concerned the parameters of sleep and cardiorespiratory function that needed to be measured and the various factors, such as patient education or technician support, that would be needed to ensure home testing is done correctly in the home.

General Overview

Types of Devices

As described in Chapter 1, diagnostic testing for OSA can be done through techniques and devices classified into four types. The most comprehensive is designated Type 1: attended, in-facility PSG. Portable devices for home use are then classified (in descending order of complexity of elements measured, with particular emphasis on respiratory effort) as Type 2, 3, or 4.

Of the 12 studies reviewed, four dealt with Type 3 portable testing devices;[9,10,15,16] eight concerned Type 4 devices.[11-14,17-20] Most dealt with devices with similar technology that had been reviewed previously.

Only one significantly different technology that was not in the previous review ─ peripheral arterial tonometry (PAT, using the Watch_PAT 100 device manufactured by Itamar Medical Ltd., located in

Haifa, Israel and Boston, Massachusetts) ─ was reported in studies that met our inclusion criteria.[13,14,17] This is a Type 4 device because it does not measure respiratory effort or airflow.

Patient Populations

Most patient populations were recruited as consecutive subjects, typically having been referred to sleep clinics or laboratories for suspected OSA. Two studies used healthy volunteers as well,[14,17] and one referred to a group at "low risk" of apnea.[20]

More subjects were male than female; most were middle-aged (mean ages in their 40s and 50s). Studies rarely reported ethnicity or race. Of the studies reporting the data, the average body mass index measures were in the range of overweight or obese; only three studies reported comorbidities (cardiovascular disease or chronic obstructive pulmonary disease; one reported smoking behaviors).

Location of Studies

Ten studies were done outside the United States (some in multiple countries): Spain, three studies; Israel, three studies; and Scotland, Germany, Belgium, and Italy, all one study each. U.S. sites were in California, Massachusetts, and Rhode Island.

Of the included 12 studies, five investigated the use of portable devices in the patients' homes;[11,14-16,19] of these, three investigations of the portable device were also done in laboratory settings. All other studies of the portable devices were conducted solely in sleep laboratories.

Implementation of Testing Studies

Of the 10 studies in which some form of testing of the portable device occurred in a laboratory setting, the investigators did not report whether the portable device was attended or unattended for six studies; one was reported as unattended[10] and one was reported as attended but the technician was not permitted to view signals from the portable device.[9] None of the five home-based studies or substudies was attended. Thus, this body of new evidence does reflect a set of investigations (unattended portable devices) relevant to the main key question.

All studies compared portable device results against those of PSG; that is, all included studies had to have had PSG tests performed. Ten studies did these two tests simultaneously ("concomitantly," "synchronously") in the sleep laboratories; two had PSG testing only at times with variable intervals from that of the portable device test.

Of the five studies with home-based testing, the temporal relationship of it to the PSG test varied considerably. For the two with only home-based

testing, the relationship in one was between 2 and 40 days after the PSG[15] and in the other up to 30 days before a PSG.[11]

The mode of hook-ups was not reported for six studies; for an additional two studies, hook-ups could be assumed from information in the articles to have been by technicians. For the five home studies, hook-ups were only by patients in all except one study in which half of the hook-ups were by technicians and half by the patients.[11]

Education of patients for using the portable devices at home varied considerably. Two home studies did not describe these procedures; in the others, patients received a 20-minute educational intervention by a trained technician and written instructions,[15] a 15- to 20-minute training period and written instructions,[11] and no direct training or instructions but rather a "Quick Guide," instructional video, and 24-hour help-line telephone number.[16]

Scoring and Interpreting Portable Monitor Results

Portable device results could be scored automatically or manually (or both). Two studies did not describe scoring.[19,20] Scoring was described only as "automated" for five studies (four in-laboratory studies[12,13,18,21] and one in-home study[16]). In addition, three studies noted automated scoring with various manual adjuncts or corrections;[9-11] one specified automated

scoring with a proprietary algorithm,[14] and one used manual and automated scoring of apnea and hypopnea and reported that automated scoring did not correlate well with the manual scoring of the PSG results.[15]

Virtually no studies described the qualifications of the individuals who interpreted the results of portable monitors. In one study, an "experienced neurophysiologist" interpreted results;[10] in another, a physician was the investigator responsible for portable device interpretation.[11]

Sponsorship of Studies

Several studies were done by the same teams of investigators. For six studies, the sources of funding were not given (although in some cases the apparent source of funding was the device manufacturer). In several cases, some or all of the investigators were employees of or consultants or scientific advisers to the company manufacturing the device.[12-14,17]

Quality of Studies and Strength of Evidence

Of these 12 studies, seven provided "Level I" evidence: random or consecutive assignment to the study as a whole; outcomes evaluators blinded to the results of the PSG or the portable device results; and reference standard (PSG) done on both populations. Two studies were "Level II," which blinded outcomes evaluations and did reference standards but did not assign patients randomly or consecutively.[10,19] Evidence Level

III (outcomes evaluators blinded and patients assigned randomly or consecutively, but no reference standard performed on all patients) was not a relevant category for this review because we required a PSG on all subjects as an inclusion criterion. Three studies (all in-laboratory) were Level IV evidence (no blinding[12,13,17]).

Using these levels of evidence categories and scores on the eight quality indicators (present or absent), we graded the quality of studies as good, fair, or poor (see Table 2 in Chapter 2). When data were not reported, we assumed the quality indicators were not met (i.e., were absent). Of these 12 studies, one was considered good;[15] four were considered fair;[10,11,14,18] and seven were considered poor.[9,12,13,16,17,19,20]

Generally, the overall quality grades for this set of studies could be characterized as at best only fair; we did not drop studies from inclusion in the report itself because of a quality grade of poor. Most studies were small to medium sized (except for the one large study of commercial drivers).

Important Findings from Reviewed Studies

For this update, we focused on the five studies with in-home results (regardless of whether those studies also included in-laboratory results). However, four studies with in-laboratory results comparing portable devices

with PSG results were rated good or fair, and we briefly comment on those just below. Authors variably reported sensitivity and specificity statistics and/or likelihood ratios (LRs); the latter provide a convenient way to estimate the change in the probabilities that a disease is present after test results are known relative to the pretest probabilities.

Portable Devices Studied in the Laboratory

Three teams compared Type 3 devices against PSG in the laboratory setting (Evidence Tables 1, 2, and 4).[10,15,16] In one study, the investigators used simultaneous unattended studies and did not report on hook-up, but did comment that the outcomes interpreter was an experienced neurophysiologist.[10] For comparisons of PSG with manual scoring of the portable device at AHI $\geq$ 15 (15.8 for the portable device), sensitivity was 90.6 percent and specificity 80.8 percent; values at other AHI cutoffs were generally similar. The authors noted that manual scoring discriminated better than did automated scoring at all AHI cutoffs and that differences were greater at higher AHI cutoff points (suggesting that the portable device might underestimate severe OSA cases). In the other laboratory-based study (fair quality), the timing was simultaneous, hook-ups were done by trained technicians (and presumably the test was attended).[15] Bland-Altman plots indicated that the agreement between manual analysis

of portable and PSG recordings had a kappa statistic of 0.62 ($P < 0.001$) and agreement on nasal pressure changes a kappa statistic of 0.57 ($P < 0.01$). The authors reported that agreement between the PSG and automated portable analysis was poor.

In the Reichert study, the laboratory comparison was done simultaneously with the PSG. The hook-up for the portable device was not described, but it was likely done by trained technicians. No discussion was given of whether the technicians intervened to fix any problems encountered with the portable monitoring equipment during this attended study.[16] Using a threshold of AHI of 15 on PSG to define the presence of OSA, the sensitivity and specificity for portable monitoring was good at 95%±5% and 91%±6% respectively. The overall agreement (0.932) had a kappa coefficient for agreement beyond chance of 0.864, which is very good.

Two studies judged fair examined the performance of Type 4 devices in laboratories (Evidence Tables 6 and 12).[14,18] One study mainly studied portable devices vis-à-vis PSG in a laboratory setting (and did a small subsample in-home study reported below).[14] The in-laboratory tests were simultaneous, but the investigators did not report how the hook-ups were done in the laboratory, whether the in-laboratory tests were attended, or the

qualifications of the individuals doing the test interpretations. Portable results were scored with an automated (proprietary) algorithm; PSG results were done with manual scoring. The authors reported a correlation of 0.88 ($P < 0.0001$) for AHI/RDI results between the portable and the PSG results. Another research team did simultaneous testing of the portable device and PSG and implied that patient hook-up in both cases was done by a technician; they did not specify if the tests were attended.[18] Using a measure of oxygen saturation and heart rate, they reported a sensitivity of 94 percent and specificity of 82 percent, for LR+ of 5.35 and LR- of 0.07. The authors concluded that if the portable test were negative, it was unlikely that the patient would receive a diagnosis of OSA.

Portable Devices Studied in the Home

As noted, our focus was on results from in-home uses of these portable devices, not in-laboratory applications. Table 4 summarizes information for these five articles; two dealt with Type 3 devices and three with Type 4 devices. We describe all five studies below and in Evidence Tables 2, 4, 6, 7, and 9.

Reichert et al. enrolled 51 patients using the NovaSomQSG (Type 3) device estimating oro-nasal airflow by sound, oxygen saturation, heart rate, respiratory effort and snoring (quality grade, poor).[16] Of these, 45

completed the study; three did not use the portable system at home and three had a portable device with a faulty memory chip. Thus, the level of missing data (i.e., patients for whom data could not be collected or analyzed of all patients entered into the study) was 13 percent. The home studies were done within 7 days before or after an in-facility PSG (half before, half after). The proportion of males in the analysis group was not reported but 38 (75 percent) of the total group of 51 in the initial sample were men. The prevalence of OSA by PSG, defined as an apnea/hypopnea index (AHI) level of >15, was 47 percent (20 of 44 patients); however, the authors reported that 40 of the patients had "split-night" studies in the laboratory, suggesting that a larger percentage had some degree of OSA. AHI scores for the home study were averaged across 3 nights. Using this average would tend to decrease the effect of variability of results of the home device and thus tend to increase the observed agreement with PSG results.

When the AHI defining a positive test was set at 15 or higher, the sensitivity for the home tests was 91 percent (± 6 percent) and the specificity was 83 percent (± 8 percent); these values led to a positive predictive value of 83 percent and a negative predictive value of 91 percent. The sensitivity/specificity values give an LR+ of 5.35 and an LR-

of 0.11, indicating a modest increase in probability of OSA if the test is positive and a modest decrease in the probability if it is negative. A Bland-Altman plot showed wide confidence limits for agreement on AHI with 2 SD approximately equal to 60.

Dingli et al. enrolled 61 patients with Type 3 equipment (Embletta) that measured flow by nasal pressure detector, thoraco-abdominal movement, pulse oximetry, and body position (quality grade, good).[15] Overall, 18 percent of these home studies had inadequate data, leaving 50 patients in the analysis. The rate of data loss was reported to improve because of a "learning effect" among the last two-thirds of patients studied, but the rate of loss was still 12 percent even in these patients. The overall prevalence of sleep apnea defined by a PSG AHI greater than 15 was 76 percent.

Using two different thresholds for the number of apneas plus hypopneas per hour (A+H/hr) in the home study to classify patients as positive or negative for OSA (home A+H/hr > 20 = positive; home A+H/hr < 10 = negative), no false-positive and no false-negative cases occurred. However, 36 percent of cases had home study results with values for A+H/hr in the "indeterminate" range; thus, the investigators could not classify these patients as having or not having OSA. The authors reported

that agreement between the Embletta automated scoring and PSG results yielded a kappa coefficient of 0.10 (poor agreement) and agreement between Embletta manual scoring and PSG a kappa of 0.54. The A+H/hr measured on the portable study agreed more closely with that from the PSG when thoraco-abdominal movement was used in conjunction with flow by nasal pressure to determine hypopneas. The authors stated that use of nasal pressure as the sole indicator of hypopneas would result in unscorable traces in approximately 8 percent of studies and recommended use of both measures to define hypopneas. In summary, in this study, although no misclassification occurred using different thresholds for a positive or a negative home study, a large proportion of patients would require additional testing.

Golpe et al. enrolled 55 patients in their home study using a Type 4 device (Apnoescreen) that recorded body position, wrist actimetry, pulse oximetry, pulse rate, and oronasal airflow by thermistry (quality grade, fair).[11] Overall, 20 percent of the patients in the home studies produced no interpretable data – 7 percent of 28 patients whose equipment was set up by a technician and 33 percent of 27 patients who did their own hook-ups.

For detection of OSA defined as an AHI of 10 or greater by PSG, the best sensitivity and specificity estimates from the “knee” of the receiver

operating characteristics (ROC) curve were approximately 90 percent and 80 percent, respectively. These values yield a LR+ of 4.5 (considered a modest change in likelihood of OSA from the pretest probability if the test is positive) and a LR- of 0.125 (considered a modest decrease in the likelihood of OSA if the test is negative).

These authors also reported the agreement between therapeutic decisions derived from the home studies and PSGs. The clinical decision on whether to treat the patient with CPAP based on the home study result agreed with that based on the PSG in 34 of 44 cases (77 percent). In six of the 10 cases of nonagreement, diagnostic classifications clearly differed: one home study false-negative result and three false-positive results. In three cases, although both types of study classified the patients as having OSA, the home study overestimated the severity inasmuch as the clinician judged that only conservative treatment was indicated. In the remaining case, the home study was inconclusive but PSG results showed that CPAP was indicated.

Liesching et al. retrospectively studied at total of 36 patients with a Type 4 device (SNAP) that monitored oronasal sound to estimate airflow (quality grade, poor).[19] Four patients had only split-night PSGs (with CPAP initiated during the study), so the investigators could not compare their

results with SNAP home study results. SNAP results were also inadequate to estimate AHI in one additional patient, leaving 31 patients in the analysis (for a 3 percent rate of missing data).

Patients were classified by PSG results as normal (AHI < 5), mild OSA (AHI 5 -15), moderate OSA (AHI 15 - 30), or severe OSA (AHI > 30). Of the eight patients who were normal by PSG, SNAP classified all as having OSA (six mild, one moderate, one severe); these figures yielded a specificity of 0 percent. Considering only the normal patients whom SNAP classified as either moderate or severe OSA as false-positives, specificity was 75 percent. SNAP results classified 21 of the 23 OSA cases (by PSG) as positive (91 percent sensitivity); one false-negative case by SNAP was classified as severe by PSG and the other as moderate.

The LR+ is 0.91 if all false-positive cases are included in specificity; it is 3.64 if only those normal patients misclassified as moderate or severe OSA by SNAP are considered. The LR- value is 0.12 in this latter circumstance. Thus, the test provides little change in the estimated probability of OSA if it is positive and a modest reduction in the probability of OSA if it is negative.

Bar et al. used a new Type 4 device measuring peripheral arterial tonometry, pulse oximetry, and actigraphy (Watch PAT 100) in home

studies in 14 patients (quality grade, fair).[14] This study had both a simultaneous in-laboratory comparison with PSG (see above) and a "subset" of 14 patients studied in the home; on two additional nights of the latter, three studies did not produce usable data, for a data loss of 11 percent.

Sensitivity and specificity were not explicitly reported for the home study. However, from the RDI values from home and PSG testing given for each subject (Figure 5, p. 699) and defining the presence of OSA by a PSG RDI of 20 or higher, we calculated that nine patients had OSA (64 percent prevalence). PAT home study detected seven of these nine cases (sensitivity, 78 percent) and had two false-positive results (specificity, 60 percent). One false-positive result by PAT had a PSG RDI of approximately 19, so if this case is considered correctly classified as positive by PAT, then sensitivity rises to 80 percent and specificity to 75 percent. Even with these estimated values of sensitivity and specificity, LR+ is 3.2 and LR- is 0.27, both indicating little effect of test results on the estimated likelihood of OSA.

Table 3. Twelve Included Studies with Type of Device, Site of Test, and Quality Grade

Study	Type and Name of Device	Site of Tests	Quality Grade (Evidence Table*)
Calleja et al., 2002[10]	Type 3 MERLIN	Sleep laboratory only	Fair (ET1)
Marrone et al., 2001[9]	Type 3 POLYMESAM	Sleep laboratory only	Poor (ET3)
Dingli et al., 2003[15]	Type 3 Embletta	Home and sleep laboratory	Good (both settings) (ET2)
Reichert et al., 2003[16]	Type 3 NovaSom QSG	Home and sleep laboratory	Fair (lab study) Poor (home study) (ET4)
Ayas et al., 2003[13]	Type 4 Watch PAT100	Sleep laboratory only	Poor (ET5)
Pillar et al., 2003[17]	Type 4 Watch PAT100	Sleep laboratory only	Poor (ET 10)
Shochat et al., 2002[12]	Type 4 SleepStrip™	Sleep laboratory only	Poor (ET11)
Zamarron et al., 2003[18]	Type 4 Criticare 504 oximeter	Sleep laboratory only	Fair (ET12)
Gurubhagavatula et al., 2004[20]	Type 4 Not reported	Sleep laboratory only	Poor (ET8)
Bar et al., 2003[14]	Type 4 WatchPAT 100	Home (substudy) and sleep laboratory	Fair (both settings) (ET6)
Golpe et al., 2002[11]	Type 4 Apnoescreen-I	Home only	Fair (ET7)
Liesching et al., 2004[19]	Type 4 SNAP technology	Home only	Poor (ET9)

*Evidence tables are found in Appendix C.

Table 4. Summary of Studies of Portable Monitoring Devices Conducted in the Home

Study and Country	Device Tested (Type)	Patient Character-istics	Timing of Home Testing	Number of Patients Enrolled / Completed (Percentage with Missing Data)	Sensitivity and Specificity; kappa*	Other Comments; Quality Grade†
Reichert et al., 2003[16] USA	NovaSom QSG (Type 3)	OSA suspected Elderly: Age range 30–83 years	Half done 7 days before and 7 days after PSG	51/45 (13%)	RDI ≥15: Sens: 91% Spec: 83% k: NR	PPV: 83% NPV: 91% Quality: II-D (poor)
Dingli et al., 2003[15] Scotland	Embletta (Type 3)	OSA suspected Elderly: Mean age 46 ± 10 years	On 2 separate nights in random order, in 2 to 4 days	61/50 (18%)	NR k: 0.54	Thresholds for OSA set at 20 but compared portable at this threshold and used 15 for the PSG Quality: I-A (good)

Table 4. Summary of Studies of Portable Monitoring Devices Conducted in the Home (cont'd)

Study and Country	Device Tested (Type)	Patient Character-istics	Timing of Home Testing	Number of Patients Enrolled / Completed (Percentage with Missing Data)	Sensitivity and Specificity; kappa*	Other Comments; Quality Grade†
Lieshing et al., 2004[19] USA	SNAP (Type 4)	OSA suspected Elderly: Mean age 50.3, range 29 to 77 years	Mean followup of 5 months (range of 2 to 10 months)	36/31 (16%)	NR Estimated from the article for patients with AHQ ≥5: Sens: 91% Spec: 0% k: 0.23	Accurately assessed severity: 12 of 31 patients (39%); Accurately predicted AHI: 17 of 31 patients (55%) within 10 events per hour Quality: II-D (poor)

Table 4. Summary of Studies of Portable Monitoring Devices Conducted in the Home (cont'd)

Study and Country	Device Tested (Type)	Patient Character-istics	Timing of Home Testing	Number of Patients Enrolled / Completed (Percentage with Missing Data)	Sensitivity and Specificity; kappa*	Other Comments; Quality Grade†
Golpe et al., 2002[11] Spain	Apnoe-screen-I (Type 4)	OSA suspected Elderly: Mean age 52.7 ± 13.3 years	Portable done first, PSG done within 30 days	55/44 (20%)	NR: Estimated from ROC curve for AHI ≥10: Sens: 90% Spec: 80% k: 0.734	Half of patients were hooked up by technicians, half by patients themselves Quality: I-C (fair)

Table 4. Summary of Studies of Portable Monitoring Devices Conducted in the Home (cont'd)

Study and Country	Device Tested (Type)	Patient Character-istics	Timing of Home Testing	Number of Patients Enrolled / Completed (Percentage with Missing Data)	Sensitivity and Specificity; kappa*	Other Comments; Quality Grade†
Bar et al., 2003[14] Israel	Watch PAT100 (Type 4)	OSA suspected Elderly: Mean age 41.4 ±15.2 years	Substudy done on two additional nights	14/3 (21%)	NR; Estimated from the article RDI ≥ 20: Sens: 78% Spec: 60% k: NR	Did not report home study results versus in-lab PSG results; Correlation of in-lab portable results with PSG: r = 0.88 (P < 0.0001) Quality: II-C (fair)

AHI: apnea/hypopnea index; NR: Not reported; OSA: obstructive sleep apnea; PSG, polysomnography; PPV, positive predictive value; NPV, negative predictive value; RDI, respiratory disturbance index

*kappa (k): level of agreement between polysomnography categorization and categorization by portable device tested.

**Percentage of missing data calculated as number of patients not analyzed because of missing data for any reason as a percentage of the total number enrolled.

†Quality grade: see Chapter 2 and Table 2 for explanation

Chapter 4. Discussion

Introduction

In interpreting these findings from our updated systematic review of the effectiveness of portable monitoring devices, as judged against that of in-facility polysomnography (PSG) for diagnosing obstructive sleep apnea (OSA), we call attention to three important questions. *First,* do the results of the more recently published studies using similar monitoring technologies differ significantly from those of the studies in the previous review done for the American Academy of Sleep Medicine (ASSM), the American College of Chest Physicians (ACCP), and the American Thoracic Society (ATS).[3-5] Specifically, do recent studies of the use of portable monitors *in the home* indicate that accuracy in diagnosing OSA is better, worse, or unchanged from the levels of accuracy reported in previously published studies? *Second*, do new portable monitor technologies demonstrate significantly different effectiveness in accurately detecting OSA? *Third,* do these recently published studies have important limitations affecting either internal or external validity?

In updating our original evidence report, we presented in Chapter 3 results and findings from 12 studies that met our inclusion criteria. These

12 studies, which we fully reviewed, fell into four categories: Type 3 devices used in laboratory settings (four studies[9,10,15,16]); Type 3 devices tested in homes whether or not they were also tested in facilities (two studies[15,16]); Type 4 devices in laboratory settings (six studies[12-14,17,18,20]); and Type 4 devices tested in homes (whether or not in facilities, three studies[11,14,19]). We focused on the five studies with in-home testing, because the questions we were asked concerned the effectiveness of unattended monitoring in the home.

We discuss here the collective knowledge base from this newer work in the context of what was known after the previous review, drawing attention to the critical issues in evaluating portable monitors generally and the issues of evaluating performance against the reference standard (PSG). These are essentially issues of the internal validity of the studies we included. In addition, we consider factors that may affect the clinical usefulness of portable monitoring for OSA in the Medicare patient population; this is a question of the external validity or generalizability of these studies to the population of interest to CMS.

Critical Issues in Evaluating Portable Monitor Studies

The effectiveness of home portable monitors is judged chiefly in terms of how well they correctly identify patients with and without clinically

significant OSA. "Clinically significant" turns on issues of severity, measured by (for instance) an apnea/hypopnea index (AHI) or a respiratory disturbance index (RDI). "Correct identification" rests on whether the likelihood of a correct diagnosis or classification (as to the presence or absence of OSA) is better after the portable device test than it is before the test. Assessing this body of evidence requires appreciation of several limitations of published studies, as these problems place a ceiling on the level of internal validity (i.e., the extent to which these studies are free of systematic bias).

Portable Testing in Laboratories or in Homes

Most articles provided only comparisons of the results from portable monitoring done simultaneously with full PSG in the laboratory, i.e., "a side-by-side" study. Although this type of study does control for night-to-night variability in the important AHI (a measure of severity) observed by PSG, it does not provide information on the performance of the equipment unattended in patients' homes where usually no technical support is available (except, perhaps via a telephone help line).

Data Loss

Data loss in this context means that some or all portable monitoring measures for individual patients originally entered into the studies were not

recorded in usable form, meaning that those patients had to be excluded from some or all analyses. Reported data loss in the home studies considered for this update ranged from 3 percent to 33 percent (in a subgroup). Moreover, at the upper end of this data loss range, many experts doing systematic reviews of clinical literature would probably regard the studies as being of only poor quality and perhaps not give them further consideration.

Only one home study *directly* compared the data loss rate between hook-up for the portable equipment by technicians and that by patients;[11] the investigators reported a 7-percent loss for technician hook-up and 33-percent loss for patient hook-up. In the study using PAT technology,[14] three of 28 (11 percent) of initial home studies set up by the patients were "rejected" whereas only three of 102 (3 percent) of studies done in the laboratory with equipment hooked up by a technician were "rejected."

Thus, although only a limited amount of evidence in the reports reviewed addresses this issue, data loss appears to be greater when the patient performs the hook-up of the equipment.

Manual or Automated Scoring

Manual *versus* automated scoring remains a significant question. Some portable monitors report a score for respiratory disturbance derived

from an automated scoring algorithm. Others provide data in a format which was later scored manually by a technician or physician.

Four studies provided insights into this issue. Reported agreement between the PSG value of apneas plus hypopneas per hour of time in bed and that derived from the portable equipment was better for manual than automated scoring for Embletta data.[15] The kappa statistic is a measure of agreement between two results beyond chance: the larger the value the better the agreement. In this study, kappa was 0.62 for manual scoring in the studies done simultaneously in the laboratory and 0.54 for those done in the home studies. The kappa statistics for the automated scoring were 0.28 and 0.10 indicating a poor agreement beyond chance. In another study, the investigators did not compare results for automated and manual scoring directly but did report area under the curve (AUC) for ROC curves derived from each method of scoring the portable data.[11] The larger the AUC, the better the performance of the test. In this study, the AUC was 0.89 for manual scoring and 0.86 for automated scoring. A third study provided a Bland-Altman plot of agreement between AHI from the PSG and that from automated scoring of the data from the NovaSom portable equipment.[16] Although the mean difference appeared to be small, the limits of agreement estimated by $\pm$ 2 standard deviations (SD) were $\pm$ 60 events

per hour. Finally, the differences in AHI between the PSG and that for the Merlin portable equipment scored manually or with an automated method were reported in a simultaneous in laboratory study.[10] The mean difference was -4 ±14 for the manual scoring and -24 ± 30 for the automated scoring.

In short, results for automated methods of scoring respiratory events appear to provide less agreement with PSG results than do manual methods. That is, for portable monitors with data recordings that could be scored either manually or with automated algorithms (or both), manual scoring produced results with better concordance with PSG results. Several studies, however, apparently used only automated scoring of the portable device results, and one used a proprietary automatic scoring system.

As a related matter, all studies apparently used experienced technicians or physicians to interpret the PSG and the portable monitoring data when the latter could be scored manually, although not all studies reported on interpreter qualifications directly. This level of experience is critical to allow adequate detection of artifacts or situations in which patient data may be questionable.

Core Aspects of Clinical Management

Appropriate management of patients with sleep-related breathing disorders requires consideration of the clinical and physiologic consequences of sleep-disordered breathing, not just a classification of severity by AHI or RDI. The overall clinical diagnosis of OSA ─ and even more importantly the decision on the appropriate management of OSA when it is present ─ depends on additional factors. Therefore, evidence that an in-home portable test can measure an RDI as well as a PSG can in the same patient is not adequate in and of itself to evaluate the usefulness of that test clinically.

The medical history and examination supply crucial information in this regard, but also very important is the apparent impact of sleep-disordered breathing on sleep quality. This can include, for example, the amount of slow wave or “deep” sleep and rapid eye movement (REM) sleep and the frequency of brief arousals and full awakenings. The portable monitors in this review did not include electroencephalogram (EEG) or electro-oculogram (EOG) signals, so investigators could not perform sleep staging or score EEG arousals or awakenings associated with respiratory or other events.

Some portable monitors use other methods for estimating sleep (e.g., actigraphy). Although these measures may have significant overall correlation with total EEG sleep during the PSG, they cannot be used to stage sleep or to detect brief cortical arousals ("micro" arousals) associated with respiratory or other events.

Other Clinical Issues

Spontaneous arousals may have an important impact on sleep quality. They can be correctly identified by manual scoring of a full PSG, but they may be missed or possibly inappropriately scored as primary respiratory-related events by automated portable scoring algorithms. The latter situation could arise if such arousals changed ventilation, for instance by increasing ventilation and oxygen saturation (SaO_2) from the typically lower baseline ("asleep") values to the normally higher "awake" values followed by a reduction in ventilation and "desaturation" to the sleeping baseline values again. This is especially problematic in patients with underlying heart or lung disease with relatively low baseline SaO_2 levels.

Other conditions producing arousals, such as Restless Legs Syndrome/Periodic Limb Movements of Sleep, are not detectable by portable monitors without electromyography (EMG) signals. These arousals may also be associated with changes in SaO_2 and misinterpreted

as primary respiratory-related events as described for spontaneous arousals.

If REM sleep is not appropriately detected, the clinicians' ability to assess the clinical impact of sleep-disordered breathing is reduced. For many patients with OSA, the severity of sleep-disordered breathing is much greater during REM sleep. The overall AHI may indicate a lower severity than clinicians might consider clinically pertinent, especially if the amount of REM sleep during the study is lower than the amount usually experienced during sleep at home unaffected by monitoring.

Likewise, nonobstructive hypoventilation events with oxygen desaturations are common during REM sleep, especially in persons with underlying lung disease, obesity, or neuromuscular weakness. Without information on sleep stage and respiratory effort, these events may be misinterpreted as obstructive events consistent with OSA. Conversely, using time in bed or recording time rather than EEG-documented sleep time to calculate the respiratory disturbance index (RDI) may produce a spuriously low value if the patient has significant time awake during the study. Sleep efficiency is a measure of the amount of time the person was asleep during the testing, for example, from the time the patient was to go to sleep, "lights out," to the awakening time "lights on." Of the studies that

took place in the laboratory only (side by side comparisons), there were two studies that reported sleep efficiencies ranging from 65.0 (± 20.9%, indicating a wide variability among the patients) to 76 (± 2%, standard error from the mean). One study reported that their home study sample had a mean sleep efficiency, as measured in the laboratory, of 82 (± 1% standard error from the mean). Thus, if the actual time were not known, the time used to calculate the time in bed could be 25% to 35% longer than actual sleep time.

Co-existing Conditions

Comorbid conditions can have a significant impact on sleep and sleep-related respiratory abnormalities. Patients with underlying lung or heart disease are more likely to show significant oxygen desaturations with nonobstructive hypopneas. Also, periodic breathing with a central apnea component (Cheyne-Stokes breathing) is common in patients with significant heart failure or atrial fibrillation; if adequate measures of respiratory effort are not available, then these conditions may be mistaken for OSA.

Generally, the studies we reported here said little about coexisting conditions in these patient populations. One group noted that they had

excluded patients using oxygen, those with certain current medications, and those who were "medically unstable."[13]

No in-home study gave information on whether comorbid conditions appeared to affect the rate of false-positive or false-negative cases. In sum, evidence is insufficient to draw conclusions about the effect of comorbidity on the effectiveness of portable devices as compared with that for PSG.

Similarity of Update and Prior Findings

The in-laboratory simultaneous studies in this review, which used technologies identical or similar to those in the studies reviewed by us for the AASM/ACCP/ATS, produced sensitivity and specificity results for diagnosing OSA similar to those previously reported. That is, the newer studies produced no meaningful changes in the level or quality of evidence for the effectiveness for home monitoring devices in diagnosing OSA. Chapter 1 summarized the earlier findings and conclusions.

Three in-laboratory studies in this review used portable monitoring technologies identical or similar to those in the 2002 review;[9,10,12] two produced sensitivity and specificity results for diagnosing OSA that were similar to those previously reported, and one reported slightly better results. We found no significant overall change in the level or quality of evidence for

the effectiveness for home monitoring devices in diagnosing OSA when used in an attended laboratory setting.

The four in-home studies that employed technologies similar or identical to those in the AASM/ACCP/ATS review had sensitivity and specificity values similar to those previously reported.[11,15,16,19] Thus, the newer studies yielded no major information that would change the previous basic conclusions about portable devices used in the home.

Three studies used a device based on a technology not considered in the prior review ─ namely, peripheral arterial tonometry (PAT). Two were done only in the laboratory setting and did not demonstrate significantly better accuracy in diagnosing OSA than other devices.[13,17] The only in-home study using PAT produced likelihood ratios indicating little effect of the test results on estimates of the probability of OSA.[14]

Issues of Internal Validity of Reviewed Studies

Factors that may falsely lower the apparent accuracy of home portable monitoring studies in detecting OSA were discussed in the original AASM/ACCP/ATS review. These include night-to-night variability, the lack of complete consensus on the definition of clinically significant respiratory events during a PSG, different sleep architecture and/or body position in the sleep laboratory different from that in the home, and the lack of clinical

validity of a single AHI or RDI threshold to define the presence or absence of OSA when comparing PSG and home study results.

These issues are all still relevant to the studies reviewed in this update. In some instances these factors may have accounted for some of the difference in results observed between portable monitoring in the home and the in-laboratory PSG, thereby causing a spurious lowering of the reported diagnostic accuracy of the home study.

One way to interpret results of tests such as this is to determine whether, relative to the pretest probability, the testing changes the post-test probability (likelihood) that a condition, in this case OSA, is present or not. This determination is based, in part, on whether the test results are positive or negative. Likelihood ratios (LRs) for a test result from the device under investigation provide a convenient way to make this determination, Positive LRs (LR+) reflect the ratio of the percentage of patients with a disease correctly identified by the test result (true positives) to the percentage of patients without the disease who are misidentified (false positives); negative LRs (LR-) reflect the ratio of the percentage of patients with a disease who have a negative test result (false negatives) to the percentage of patients without the disease who have a negative test result (true negatives).

LRs (positive or negative) of 1.0 could be said to reflect a useless test. LR+ values of 2 to 5 show modest effect of the test in ruling in a diagnosis (i.e., concluding the disease is present); those at the 5 to 10 level have strong effect. LR- values between 0.2 and 0.5 show a modest impact on ruling out a diagnosis (i.e., concluding that the disease is absent); those between 0.1 and 0.2, a moderate impact, and those less than 0.1, a strong effect. The original review and the accompanying paper on comparing diagnostic tests[4] review these points in detail in the context of OSA.

Mixed evidence (one good study, two fair, one poor) showed that Type 3 devices, when used in an attended laboratory setting, can modestly increase or modestly decrease the likelihood of an accurate OSA diagnosis relative to pretest probabilities. Somewhat more questionable evidence (2 fair, 4 poor) suggests that Type 4 devices, when used in the attended laboratory setting, can also modestly increase and decrease the likelihood of correctly determining the presence or absence of OSA. Data loss (i.e., missing data for individual patients) of 10 percent to 20 percent should be expected in home studies.

One good study and one poor quality study indicate Type 3 monitoring devices, when used in unattended home settings, can both modestly increase and modestly decrease the probability of OSA relative to

the probabilities before the testing.[15,16] LRs in this context can be improved by using different thresholds for RDI on the portable test to increase (rule in) and decrease (rule out) the probability of OSA. However a relatively high proportion of patients (up to 40 percent) may then be "unclassifiable" and need further testing.

Three studies (two fair, one poor quality) using Type 4 monitoring devices were done unattended in the home.[11,14,19] These authors reported data indicating that these devices can modestly reduce the probability of OSA (e.g., reach a LR- of 0.2 or lower).[11,19] Less evidence exists that these devices can increase the probability that OSA is present (e.g., reach or produce a LR+ of greater than 5). Overall, data loss of 3 percent to 20 percent was reported and up to 33 percent when the patient did the hook-up themselves. The only in-home study using nonstandard Type 4 monitoring technology (PAT[14]) received a fair quality rating and produced LRs indicating little effect of the test results on estimates of the probability of OSA.

Only one study addressed how clinical management decisions based on the portable test results would compare with those based on PSG results.[11] In 10 of 44 cases (23 percent), the clinical decision on whether CPAP was indicated based on the interpretation of the portable results

differed from that based on interpretation of the PSG. In six of the cases for which the management decisions differed between PSG and home study results, the home study was deemed a false-negative or a false-positive result.[11] In the remaining 4 cases, the interpretation of the home studies agreed with that of the PSG about the presence of OSA, but the severity grading differed significantly and recommended therapy differed.

These studies were done in a highly selected patient populations with high prevalence rates of OSA by PSG (50 percent to 75 percent) and proportions of males (approximately 75 percent to 90 percent). The overall prevalence of comorbid conditions in the patients studied was typically not stated, and no characterization of comorbid conditions was given for those patients who were incorrectly classified (false positives and false negatives) or were "unclassifiable" by portable testing.

External Validity or Generalizability

Overall Generalizability

The published studies of portable monitors have several limitations in regard to generalizability of their results to less highly selected patients (i.e., populations with characteristics different from those of the samples studied). Studies in this update were done on patients identified as having a high pretest probability of having OSA by PSG; prevalence of OSA by

PSG was generally 50 percent or greater, and in some cases the PSG prevalence rate approached 80 percent. Most of these studies had a majority of males and did not report the proportion of patients with significant comorbid conditions such as chronic obstructive pulmonary disease, asthma, congestive heart failure, or neuromuscular disorders.

Generalizability to the Medicare Beneficiary Population

No study specifically targeted an elderly population. Apart from that, applying findings in these studies to the Medicare population has several limitations over and above the issues raised with respect to overall generalizability and internal validity.

First, the prevalence of reported excessive daytime sleepiness in the elderly is known to be high. For example, of 4,578 noninstitutionalized Medicare enrollees, 20 percent reported being "usually sleepy in the day time."[22] The prevalence of medical conditions associated with poor quality sleep and daytime sleepiness for reasons other than OSA is higher in the elderly than in younger populations. In one sample of 18,980 subjects, the prevalence of Restless Legs Syndrome increased with age (ages 40 to 49 years, 4.7 percent; ages 60 to 69 years, 8.3 percent; ages 70 to 79 years, 8.9 percent).[23]

Second, "classical" signs and symptoms of SDB are less closely associated with OSA in the elderly than in other age groups. A study of 5,615 community-dwelling adults found that "as age increased the magnitude of associations of SDB and body habitus, snoring and breathing pauses decreased" (p. 893).[24] The authors concluded that breathing pauses and obesity may be "particularly insensitive" for identifying SDB in the elderly. Thus, the prevalence of true OSA in Medicare patients referred for sleep studies because of signs and symptoms such as obesity and excessive daytime sleepiness may be significantly different from the prevalence in the patient groups in the studies reviewed.

Finally, we see no indications that adequate sleep study data are more difficult to obtain in unattended home studies in the elderly than in other groups. The Sleep Heart Health Study reported no significant effects of age or sex on the overall success rate in obtaining interpretable data,[25] although these investigators did observe a significant decrease in the duration of adequate abdominal "effort" signals in the elderly. Thus, the proportion of interpretable data obtainable in home studies on Medicare patients is likely to be similar to that from the patients in the study populations.

However, the rates of false-positive and false-negative tests may differ because of the higher prevalence of comorbid conditions in the Medicare population.[22,23] This may also occur because the prevalence of OSA in Medicare patients who are referred for study because of excessive daytime sleepiness may be lower than the OSA prevalence in the studied patients.[1] This may also be true for patients referred by physicians without special training in sleep medicine, but no explicit information on this is available. Most reported studies derived patients from specialized sleep clinics; only one study in the review explicitly stated that patients were referred by "community physicians."[16]

Summary

The key questions posed for this review asked how portable sleep testing devices compared to PSG in diagnosing OSA and, assuming equivalent effectiveness, what sleep and physiologic factors and what patient and technician conditions were important to measure or have in place.

We updated an earlier evidence report with a systematic literature search and in-depth review of 12 articles that met inclusion criteria for addressing these questions. Most articles covered technologies (of Types 3 and 4 only) that had been examined previously; three dealt with a single

new technology. Most studies involved testing home devices against PSG in a sleep laboratory; five studies either wholly or partially examined home devices in the home.

This newer body of evidence does not materially change earlier findings regarding in-home devices for diagnosing OSA.[3] Of the five in-home studies, two[15,16] were done with Type 3 devices (one of good quality, and one of poor quality) and three[11,14,19] with Type 4 devices (two fair and one poor quality). Information from the one in-home study of a new technology, of fair quality, gave little support for concluding that it was better than any other Type 4 device.

Choices of cutoffs for determining OSA by AHI or RDI differed widely across these studies, making cross-study comparisons impossible. The better studies yielded sensitivity and specificity values (or LRs) that provided modest changes the probability of OSA over the pretest probability. In studies that directly compared automated vs. manual scoring of data from home monitoring devices, manual scoring correlated better with data from laboratory PSG.

Improved sensitivities and specificities could be achieved by using two different thresholds to define results of a home test as “positive” or “negative” for OSA, but this left a large proportion of patients with

"indeterminate" results. The clinical decision about the need for CPAP therapy based on the interpretation of the home study differed from that based on the PSG in 23 percent of cases in the one study which reported this type of comparison. No studies reported the effect of co-morbid conditions on the sensitivity or specificity of the home testing. The overall proportion of home studies with inadequate data averaged 13 percent but in one study data loss was as high as 33 percent when the patients performed the hookup compared to only 3 percent when hookup was done by technicians. Mean age for patients in the home studies ranged from 41.4 to 52.7 years. No information was presented on whether the sensitivity/specificity or the rate of data loss was associated with patient age. More evidence is needed to reach conclusions about the effect of co-morbidities, age and patient versus technician performed hookup on the overall effectiveness of home studies in diagnosing OSA compared to an in-laboratory PSG.

References

1. Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S. The occurrence of sleep-disordered breathing among middle-aged adults. *N Engl J Med.* 1993;328:1230-5.

2. Ferber R, Millman R, Coppola M, et al. Portable recording in the assessment of obstructive sleep apnea. ASDA standards of practice. *Sleep.* 1994;17:378-92.

3. Flemons WW, Littner MR, Rowley JA, 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.

4. Flemons WW, Littner MR. Measuring agreement between diagnostic devices. *Chest.* 2003;124:1535-42.

5. Chesson AL Jr, Berry RB, 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.

6. ATS/ACCP/AASM Taskforce Steering Committee. Executive summary on the systematic review and practice parameters for portable monitoring in the investigation of suspected sleep apnea in adults. *Am J Respir Crit Care Med.* 2004;169:1160-3.

7. Institute for Clinical Systems Improvement (ICSI); *Diagnosis and Treatment of Obstructive Sleep Apnea*. Bloomington (MN): Institute for Clinical Systems Improvement (ICSI); 2004.

8. Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB; *Diagnosis and Screening in Evidence Based Medicine: How to Practice and Teach EBM*. Edinburgh: Churchill Livingstone; 2000;67-93.

9. Marrone O, Salvaggio A, Insalaco G, Bonsignore MR, Bonsignore

G. Evaluation of the POLYMESAM system in the diagnosis of obstructive sleep apnea syndrome. *Monaldi Arch Chest Dis.* 2001;56:486-90.

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

11. Golpe R, Jimenez A, Carpizo R. Home sleep studies in the assessment of sleep apnea/hypopnea syndrome. *Chest.* 2002;122:1156-61.

12. Shochat T, Hadas N, Kerkhofs M, et al. The SleepStrip: an apnoea screener for the early detection of sleep apnoea syndrome. *Eur Respir J.* 2002;19:121-6.

13. Ayas NT, Pittman S, MacDonald M, White DP. Assessment of a wrist-worn device in the detection of obstructive sleep apnea. *Sleep Med.* 2003;4:435-42.

14. Bar A, Pillar G, Dvir I, Sheffy J, Schnall RP, Lavie P. Evaluation of a portable device based on peripheral arterial tone for unattended home sleep studies. *Chest.* 2003;123:695-703.

15. 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.

16. Reichert JA, Bloch DA, Cundiff E, Votteri BA. Comparison of the NovaSom QSG, a new sleep apnea home-diagnostic system, and polysomnography. *Sleep Med.* 2003;4:213-8.

17. Pillar G, Bar A, Betito M, et al. An automatic ambulatory device for detection of AASM defined arousals from sleep: the WP100. *Sleep Med.* 2003;4:207-12.

18. Zamarron C, Gude F, Barcala J, Rodriguez JR, Romero PV. Utility of oxygen saturation and heart rate spectral analysis obtained from pulse oximetric recordings in the diagnosis of sleep apnea syndrome. *Chest.* 2003;123:1567-76.

19. Liesching TN, Carlisle C, Marte A, Bonitati A, Millman RP.

Evaluation of the accuracy of SNAP technology sleep sonography in detecting obstructive sleep apnea in adults compared to standard polysomnography. *Chest.* 2004;125:886-91.

20. Gurubhagavatula I, Maislin G, Nkwuo JE, Pack AI. Occupational screening for obstructive sleep apnea in commercial drivers. 2004;170: 4:371-6.

21. Pillar G, Bar A, Shlitner A, Schnall R, Shefy J, Lavie P. Autonomic arousal index: an automated detection based on peripheral arterial tonometry. *Sleep.* 2002;25:543-9.

22. Whitney CW, Enright PL, Newman AB, Bonekat W, Foley D, Quan SF. Correlates of daytime sleepiness in 4578 elderly persons: the Cardiovascular Health Study. *Sleep.* 1998;21:27-36.

23. Ohayon MM, Roth T. Prevalence of restless legs syndrome and periodic limb movement disorder in the general population. *J Psychosom Res.* 2002;53:547-54.

24. Young T, Shahar E, Nieto FJ, et al. Predictors of sleep-disordered breathing in community-dwelling adults: the Sleep Heart Health Study. *Arch Intern Med.* 2002;162:893-900.

25. Kapur VK, Rapoport DM, Sanders MH, et al. Rates of sensor loss in unattended homy polysomnography: The influence of sage, gender, obesity, and sleep-disordered breathing. *Sleep.* 2000;23:1-7.

Appendices

Appendix A.
Acknowledgments

Appendix A
Acknowledgments

We thank Elise Berliner, Ph.D., Task Order Officer for the Agency for Healthcare Research and Quality for her encouragement and support throughout this project. We extend our appreciation to Loraine Monroe, the RTI-UNC Evidence-based Practice Center document preparation specialist; to Tammeka Swinson, B.A., the EPC project manager, for considerable assistance with the literature search and abstracting tasks; and to Mark Howell, RTI librarian, for help with all searches.

Appendix B.
Data Abstraction Form

Placeholder

Appendix C
Evidence Tables

Glossary of Terms

AHI	Apnea/Hypopnea Index
AUC	Area Under the Curve
Avg	Average
BMI	Body Mass Index
CI	Confidence Interval
COPD	Chronic Obstructive Pulmonary Disease
Corr	Correlation
CPAP	Continuous Positive Airflow Pressure
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyogram
EOG	Electro-oculogram
ESRD	End Stage Renal Disease
G	Group
h	hour
H	Home part of the study
L	Laboratory part of the study
n	number
NA	Not applicable
NR	Not Reported
OSA	Obstructive Sleep Apnea (Obstructive Sleep Apnea Syndrome (OSAS) and Sleep Apnea Syndrome (SAS) are all included under OSA)
OSAS	Obstructive Sleep Apnea Syndrome
P	Probability
PAT	Peripheral Arterial Tone
PLM	Periodic Limb Movement
PSG	Polysomnography
r	Rho
RDI	Respiratory Disturbance Index
ROC	Receiver Operator Curve
s	second(s)
SAHS	Suspected Apnea/Hypopnea Syndrome
SaO_2	Saturated Oxygen
SAS	Sleep Apnea Syndrome
SD	Standard Deviation

Sens	Sensitivity
Spec	Specificity
vs	versus

Evidence Table 1. Calleja et al., 2002: Study Design

Study	Portable Device/ Implementation	Sample Selection	Study Subjects n (% of Entered)
Reference: Calleja JM, Esnaola S, Rubio R, Duran J Comparison of a cardiorespiratory device versus polysomnography for a diagnosis of sleep apnea. *European Respiratory Journal,* 2002[10] **Study Location/ Sponsor:** Spain/ Department of Health, Basque Government **Type:** 3	**Device:** MERLIN **Equipment & Methodology:** Cardiorespiratory polygraph **Channels:** Oronasal flow, chest and abdominal respiratory movement, tracheal sound, cardiac frequency, oxygen saturation, body position, CPAP **Flow:** Thermister **Effort:** NR **Portable Setting:** Sleep lab **Mode:** Unattended **Hook-up:** NR **Timing:** Simultaneous	**Site:** Sleep clinic **Inclusion:** Adult patients with clinical suspicion of sleep apnea referred from sleep outpatient clinic to sleep lab **Exclusion:** NR **Clinical Pretest Probability Distribution of Suspected Severity of OSA:** NR	**Eligible:** 86 **Excluded:** NR **Entered:** 86 **Dropout:** NR **Data Loss:** 7 (8%) **Analyzed:** 79

Evidence Table 1. Calleja et al., 2002: Study Design, cont'd

Demographics	PSG Equipment	PSG Interpretation	Portable Interpretation
Males: 89% **Mean BMI:** 30.1 Range: NR SD: ± 4.4 kg/m^2 **Mean Age:** 52 Range: NR SD: ± 11.1 years **% Non-White:** NR **Prevalence of OSA by PSG:** AHI>10: 81% Ave AHI: 41.8 SD: ± 27.7 **Comorbidities:** NR	**Setting:** Sleep Lab **Equipment & Methodology:** Alice 3 or Ultrasom **Channels:** EEG (C4-A1, C3-A2), EOG, tracheal sound with microphone, tibial and submental EMGs, ECG (modified V2 lead), respiratory sensors, chest and abdominal effort, oxyhemoglobin saturation by finger pulse oximeter, and body position **Flow:** Thermisters **Effort:** Belt sensors – piezo-electric gauge	**Scoring:** Manual- each 30 s epoch scored for sleep state, breathing, oxygenation and movement **Blinding:** To results of other method **Criteria for Excluding Data:** NR **Qualifications:** Neurophysiologist with broad experience in diagnosis of sleep apnea	**Scoring:** Automated with manual visual evaluation of printouts **Blinding:** To results of other method **Data Quality Categorized:** NR **Criteria for Excluding Data:** NR **Qualifications:** Experienced Neurophysiologist – different from the PSG interpreter

Evidence Table 1. Calleja et al., 2002: Study Design, cont'd

Comparison of Results	Special Statistics
Definitions: • Apnea: complete cessation of thermister signal of ≥ 10 s • Hypopnea: discernible reduction of ≥50% of the thermister signal for > 10 s accompanied by a decrease of ≥ 3% oxyhemoglobin saturation and/or an EEG arousal • AHI: PSG uses Total Sleep Time for a denominator vs per hour of polygraphic recording in the portable **AHI/RDI:** Avg PSG: 34.4 (29.2) Portable: 10.4 (18.2) for automated scoring; 30.5 (24.5) for manual scoring Manual score comparisons: For AHI ≥ 5 (6.7 cut off point for portable) • Sensitivity – 97.1 (93-100) • Specificity – 90.9 (74-100) For AHI ≥ 10 (9.8 cut off point for portable) • Sensitivity – 90.6 (83-98) • Specificity – 86.7 (69-100) For AHI ≥ 15 (15.8 cut off point for portable) • Sensitivity – 90.6 (83-98) • Specificity – 80.8 (66-96) For AHI ≥ 20 (21.1 cut off point for portable) • Sensitivity – 91.1 (83-99) • Specificity – 85.3 (73-79) For AHI ≥ 30 (27.6 cut off point for portable) • Sensitivity – 88.6 (78-99) • Specificity – 90.9 (82-99) For PSG AHI ≤ 10: • Automatic scoring in portable: 5.3±5.1 ♦ PSG: 3.1 ± 2.1 Portable: • Manual scoring in portable: 36.4±23.4 ♦ PSG: 41.8 ± 27.7 • Sleep Efficiency: ♦ PSG: for AHI< 10: 79.3 (12.7); for AHI ≥ 10 80.6 (14.4); total: 80.4 (14) ♦ Portable: NR	**Bland-Altman Plot:** • Difference between manual and automated analysis of the portable only • Automatic: -23.98 ± 30.2 (distribution of points is straight downward) • Manual: -3.98 ± 14 (distribution of points toward the horizontal line) **ROC:** • Discriminatory ability of the manual scores was greater than that observed with automatic scores for all AHI cutoff points (≥ 5, ≥ 10, ≥ 15, ≥ 20, ≥ 30) • Best results were achieved at ≥ 5 with manual scoring (ROC: 0.97.6) clearly superior to automatic scoring (ROC: 0.828) • Best cut off point for the portable= 6.7

Evidence Table 1. Calleja et al., 2002: Study Design, cont'd

Data Loss/ Variability/Cost Benefit	Quality Indicators	Author's Conclusions/ Limitations
Data Loss: 8% (7/86) **Reasons:** • Sleep time <240 minutes • Lack of thermister signal • Incomplete recording due to technical problems for storing data in the computer **Night to Night Variability:** NA **Cost Benefit Analyses:** NR	**Evidence Level:** • Subjects selected consecutively or randomly (no design influence): N • Both tests scored blindly: Y • Reference standard done on all subjects: Y **Quality Level:** • Prospective study: Y • Portable done outside the lab: N • Random order of allocation of subjects to PSG or Portable first: NA • Low data loss $\leq$ 10% of flow/SaO_2/EEG as applicable : Y • High % completed ($\geq$ 90% of those entered): Y • PSG methodology fully described: Y • Portable monitor methodology fully described: N • Portable scoring fully described: N **Evidence Level Score:** II **Quality Level Score:** C **Grade:** Fair	**Conclusions:** • The MERLIN device is a useful diagnostic approach for the initial assessment of adult patients with clinical suspicion of sleep apnea syndrome • Manual scoring is clearly better than automated scoring • 90 to 96% of patients were correctly classified by portable using its best cut-off points in its manual index of respiratory events • For patients incorrectly classified for low AHI cut-off points, discrepancy with PSG was small • Differences were greater with regard to high cut-off points and showed the portable had a tendency to underestimate severe sleep apnea cases **Limitations:** • Small sample, made up of mostly males with a high prevalence of sleep apnea, which limits generalizing results to a non-referral population

Evidence Table 2. Dingli et al, 2003: Study Design

Study	Portable Device/ Implementation	Sample Selection	Study Subjects n (% of Entered)
Reference: Dingli K, Coleman EL, Vennelle M et al., Evaluation of a portable device for diagnosing the sleep apnoea/hypopnoea syndrome European *Respiratory Journal,* 2003[15] **Study Location/ Sponsor:** Scotland, UK/ NR **Type:** 3	**Device:** Embletta **Equipment & Methodology:** NR **Channels:** Nasal pressure detector; thoraco-abdominal movement detection, finger pulse oximeter; body position detection **Flow:** Canulae/pressure tranducer system **Effort:** Piezoelectric belts **Portable Setting:** In-lab; Home **Mode:** In-lab: NR Home: unattended **Hook-up:** In-lab: trained technicians; Home: patients, after 20 minute education by trained technician, written instructions, and applied sensors unsupervised **Timing:** In-lab: synchronous Home trial: on two separate nights, in random order within a time interval of 2-40 days	**Site:** Sleep center **Inclusion:** Possible OSAHS **Exclusion:** Living > 50 miles from the sleep center and immobility **Clinical Pretest Probability Distribution of Suspected Severity of OSA:** NR	**Eligible:** In-lab: 40 consecutive patients Home: 61 consecutive patients **Excluded:** NR **Entered:** In-lab: 40 Home: 61 **Dropout:** NR **Data Loss:** In-lab: 1 was excluded due to technical problems as the portable only recorded 3 minutes of interpretable data In-home: 11 patients (1 due to not using equipment at all; 5 due to technical problems with the portable equipment; and 5 with no reasons reported) **Analyzed:** In-lab: 39 Home: 50

Evidence Table 2. Dingli et al, 2003: Study Design, cont'd

Demographics	PSG Euipment	PSG Interpretation	Portable Interpretation
In-lab/Home **Males:** 33/47 **Mean BMI:** 32/31 Range: NR SD: ± 6 kg/m2 **Mean Age:** 46/50 Range: NR SD: ± 10 years/± 11 years **% Non-White:** NR **Prevalence of OSA by PSG:** AHI<10: 5.3±0.9 AHI≥10<20:13.8 ±0.7 AHI≥20: 56.3±7.7 Ave AHI: • In-lab:35 h^{-1} • Home: 29 h^{-1} SD: NR PSG AHI>15: 76% **Comorbidities:** NR	**Setting:** Sleep Lab **Equipment & Methodology:** Compumedics, applied by trained technicians **Channels:** Sleep monitoring through 2 frontal unipolar, unilateral (FP1-LEOG, FP2-REOG) and bipolar sentral/occipital (CZ-PA) EEG, 2 submental EMG, bilateral tibial EMG and a body position detector; 3-lead ECG; respiration monitoring, thoraco-abdominal movement detection, a digital microphone for snoring detection, and a pulse oximeter **Flow:** thermister **Effort:** Two inductance plethysmographic belts	**Scoring:** By same observer as portable **Blinding:** To patient's identity or portable data or results **Criteria for Excluding Data:** NR **Qualifications:** NR	**Scoring:** • Both in-lab and home study were scored by same observer as PSG • Home: manual scoring of A+H by changes in nasal pressure and thoraco-abdominal movement; and manual scoring of A+H by nasal pressure alone; and automatic scoring of A+H **Blinding:** To patient's identity or PSG results **Data Quality Categorized:** NR **Criteria for Excluding Data:** NR **Qualifications:** NR

Evidence Table 2. Dingli et al, 2003: Study Design, cont'd

Comparison of Results	Special Statistics
Definitions: • Apneas: complete cessation of airflow • Hypopneas: reduction in thoraco-abdominal movement ≥ 50% for ≥ 10 s; during synchronous study, portable scored it as a reduction in the nasal pressure amplitude of ≥ 50% for ≥ 10 s • <10 (A+H)/h^{-1} in bed = Not OSAHS; 10 to 20 (A+H)/h^{-1} in bed = Possible OSAHS; ≥20 (A+H)/h^{-1} in bed = OSAHS; (manual portable scoring) **AHI:** • # of apnoeas and hypopneas AND • For PSG: After sleep onset by total sleep time • For Portable: by total time in bed (started after respiration settled down to a rhythmic stable pattern) **Apnea Index:** • Using hypopnea scoring based on nasal pressure changes alone, 1 patient with PSG (A+H >15 h^{-1}) would have been scored normal by portable **AHI/RDI:** • Avg PSG AHI: ♦ 35.4±5.5; AHI<10: 5.3±0.9; AHI≥10<20:13.8±0.7; AHI≥20: 56.3±7.7 • For AHI ≥ 15 on PSG: ♦ 21/22 of pts with OSAHS had ≥20 (A+H/h^{-1} in bed) with manual home study scoring • For AHI < 15 on PSG: ♦ 9/9 had <10 (A+H/h^{-1} in bed) with manual scoring • For AHI ≥ 15 on PSG: ♦ 23/23 had > 20 (A+H/h^{-1} in bed) ♦ Sensitivity – NR ♦ Specificity – NR • Using home A+H/hr > 20 as a positive and A+H/hr < 10 as a negative, there were no false-positives or false-negatives • 36% of home study results were in the "indeterminate range" **Sleep Efficiency:** • Simultaneous in-lab with PSG: 76 ± 2% • Home Patients as measured in lab: 82.0 ± 1%	**Bland-Altman Plot and Scatterplots:** In-lab: • manual analysis of recordings: ♦ kappa: 0.62; P<0.001; • manual analysis of nasal pressure changes: ♦ kappa: 0.57; P<0.001 Home: • manual analysis vs PSG outcomes: ♦ kappa: 0.54; P<0.001 During both the in-lab and home studies, agreement between PSG and automated portable analysis was poor (kappa: 0.28 and 0.10) **ROC:** NR

Evidence Table 2. Dingli et al, 2003: Study Design, cont'd

Data Loss/ Variability/Cost Benefit	Quality Indicators	Author's Conclusions/ Limitations
Data Loss: • In-lab: 1 due to technical problems as portable only recorded 3 minutes of interpretable data; 3 due to technical problems in the nasal pressure recordings (not recording or not interpretable) • Home: 11 (18%) (1 due to not using equipment at all, 5 due to technical problems with the portable equipment such as plugging, batteries or software, for 5 there was no reason reported); after a learning effect from the first third, the rate dropped to 12% **Night to Night Variability:** NR **Cost Benefit Analyses:** Portable reduce diagnostic costs by 42% if those in the diagnostic categories went straight to CPAP titration or no further investigation with only those in the possible OSAHS and failed home study groups proceeding to PSG	**Evidence Level:** • Subjects selected consecutively or randomly (no design influence): Y • Both tests scored blindly: Y • Reference standard done on all subjects: Y **Quality Level:** • Prospective study: Y • Portable done outside the lab: L-N; H-Y • Random order of allocation of subjects to PSG or Portable first: Y • Low data loss $\leq$ 10% of flow/SaO_2/EEG as applicable : L-Y; H-N • High % completed ($\geq$ 90% of those entered): Y • PSG methodology fully described: Y • Portable monitor methodology fully described: Y • Portable scoring fully described: L-N; H-Y **Evidence Level Score:** I **Quality Level Score:** A **Grade:** Good	**Conclusions:** • Using three diagnostic categories, 32 of 50 patients from the in-lab study were correctly categorized, but 18 of 50 fell into the "possible OSAHS" category in which further study might be indicated (of those, 15 had PSG's of ≥15) • All 34 with >15 ($A+H/h^{-1}$) on the portable had AHI's >15 • 3/16 of <15 ($A+H/h^{-1}$)on portable had AHI's <15 • The mean difference between $(A+H)/h^{-1}$ in bed was 2± 5h with a close correlation between the results of the two studies (rho=0.98, p<0.001) but with AHI the difference was 8± $16h^{-1}$ but when those with >40 per hour were excluded, the mean difference was 2± 5h • Use of nasal pressure signal as the sole indicator of hypopneas will result in unscorable traces in ~8% of studies • The automated scoring software did not relate closely to manually scored results **Limitations:** • Small number of subjects • Variance in home study could be due to night-to-night variation

Evidence Table 3. Marrone et al., 2001: Study Design

Study	Portable Device/ Implementation	Sample Selection	Study Subjects n (% of Entered)
Reference: Marrone O, Salvaggio A, Insalaco G et al. Evaluation of the POLYMESAM system in the diagnosis of obstructive sleep apnea syndrome. *Monaldi Archives for Chest Disease*, 2001[9] **Study Location/ Sponsor:** NR (assume Italy)/NR **Type:** 3	**Device:** POLYMESAM **Equipment & Methodology:** A recorder to which multiple sensors are linked for the detection of signals (listed below) **Channels:** Oxyhemoglobin saturation by finger sensor; heart rate from 3 ECG chest electrodes; snoring sound by microphone on thyroid cartilage; body posture; oronasal airflow; thoracic and abdominal movements; and limb activity or CPAP **Flow:** Three-fold thermocouple sensor for both nostrils and mouth **Effort:** Stretch belts **Portable Setting:** Sleep lab **Mode:** Attended but technician not allowed to visualize signals of portable **Hook-up:** NR, assume technician **Timing:** Simultaneous	**Site:** Sleep Lab **Inclusion:** Patients referred to sleep laboratory with suspected sleep apnea syndrome **Exclusion:** NR **Clinical Pretest Probability Distribution of Suspected Severity of OSA:** NR	**Eligible:** 50 consecutive **Excluded:** NR **Entered:** NR **Dropout:** NR **Data Loss:** NR **Analyzed:** 50

Evidence Table 3. Marrone et al., 2001: Study Design, cont'd

Demographics	PSG Equipment	PSG Interpretation	Portable Interpretation
Males: 40 **Mean BMI:** 32.7 Range: NR SD: ±6.1 kg/m^2 **Mean Age:** 49.6 Range: NR SD: ±10.2 years **% Non-White:** NR **Prevalence of OSA by PSG:** RDI>10: 84% (42/50) Ave RDI: NR SD: NR **Comorbidities:** NR	**Setting:** Sleep Lab **Equipment & Methodology:** Somnostar **Channels:** 2 EEG (C3-A2, O1-A2); right and left EOG; submental EMG; oronasal airflow; thoracic and abdominal movements; oxyhemoglobin saturation; ECG; and body posture **Flow:** Thermocouple **Effort:** Piezoelectric belts	**Scoring:** NR **Blinding:** Separate scorers blinded to paired results **Criteria for Excluding Data:** NR **Qualifications:** NR	**Scoring:** Automated, but raw data visualized and manually corrected, with the exception of ECG that can only be visualized as heart rate **Blinding:** Separate scorers blinded to paired results **Data Quality Categorized:** NR **Criteria for Excluding Data:** NR **Qualifications:** NR

Evidence Table 3. Marrone et al., 2001: Study Design, cont'd

Comparison of Results	Special Statistics
Definitions: • Central apneas: absence of airflow for at least 10 s associated with the lack of any thoraco-abdominal movement • Obstructive apneas: absence of airflow for at least 10 second, associated with the persistence of thoraco-abdominal movement • Hypopneas: discernible reductions in the airflow signal for at least 10 seconds, associated with a decrease in oxyhemoglobin saturation by at least 4% • Portable: all events normalized to per hour of time in bed vs PSG which used total sleep time • Sleep efficiency: total sleep time divided by total time in bed times 100 • OSA: AHI≥10 • AH/TIB: number of apneas and hypopneas per hour of time in bed **Apnea Hypopnea Index:** AH/TIB ≥5: • Sensitivity – 100.0 • Specificity – 71.4 • Positive predictive value: 95.5 • Negative predictive value: 100 AH/TIB ≥10: • Sensitivity – 95.2 • Specificity – 100.0 • Positive predictive value: 100.0 • Negative predictive value: 80.0 **Desaturation Index:** Lowest oxyhemoglobin saturation during sleep was 75.1±11.2%; time spent in SaO_2 below 90% was 25±24% **Sleep Efficiency:** 65 ± 20.9%	**Bland-Altman Plot:** • Good agreement between apnea/ hypopnea per time in bed; central apnea per time in bed; mixed apnea per time in bed and mean apnea/hypopnea duration • Lower obstructive apnea per time in bed was reflective of a higher hypopnea per time in bed **ROC:** NR

Evidence Table 3. Marrone et al., 2001: Study Design, cont'd

Data Loss/ Variability/Cost Benefit	Quality Indicators	Author's Conclusions/ Limitations
Data Loss: NR **Night to Night Variability:** NA **Cost Benefit Analyses:** NR	**Evidence Level Score:** • Subjects selected consecutively or randomly (no design influence): Y • Both tests scored blindly: Y • Reference standard done on all subjects: Y **Quality Level Score:** • Prospective study: N • Portable done outside the lab: N • Random order of allocation of subjects to PSG or Portable first: NA • Low data loss ≤ 10% of flow/SaO_2/EEG as applicable: N • High % completed (≥ 90% of those entered): N • PSG methodology fully described: Y • Portable monitor methodology fully described: Y • Portable scoring fully described: Y **Evidence Level Score:** I **Quality Level Score:** D **Grade:** Poor	**Conclusions:** • The number of central, obstructive, or mixed apneas, hypopneas and total number of ventilatory disorders per hour of time in bed were significantly correlated • Obstructive apnea per time in bed was lower with the portable device • Due to low sleep efficiency, all indices calculated on total sleep time were on average 35% higher than those calculated on time in bed, this resulted in significantly higher total sleep time rates with respect to total time in bed rates **Limitations:** • When using cardiorespiratory monitoring the rate of disordered breathing events are calculated per hour of time in bed and not total sleep time

Evidence Table 4. Reichert et al., 2003: Study Design

Study	Portable Device/ Implementation	Sample Selection	Study Subjects n (% of Entered)
Reference: Reichert JA, Bloch DA, Cundiff E et al. Comparison of the NovaSom QSG, a new sleep apnea home-diagnostic system and polysomn-ography. *Sleep Medicine,* 2003[16] **Study Location/ Sponsor:** California, USA/ Sequoia Hospital Pulmonary Research Fund **Type:** 3	**Device:** NovaSom QSG **Equipment & Methodology:** A bedside unit and a patient module worn on the patient's wrist and three body sensors: airflow, finger oximeter and respiration effort **Channels:** 5: nasal and oral airflow, oxygen saturation, heart rate, respiration effort and snoring sound intensity **Flow:** Nasal and oral airflow using sound **Effort:** NR **Portable Setting:** In-lab; Home **Mode:** Unattended **Hook-up:** Home: Patient picked up system from sleep lab and told to use it for 3 nights but received no instructions on how to use it but given a Quick Guide and instructional video and 24 hour help line telephone number **Timing:** • In-lab: simultaneous • Home: within 7 days before or after PSG (half of patients were done before PSG, half after)	**Site:** Sleep lab **Inclusion:** Adults suspected of sleep apnea referred to the clinic for PSG **Exclusion:** NR **Clinical Pretest Probability Distribution of Suspected Severity of OSA:** NR	**Eligible:** 51 consecutive adults **Excluded:** NR **Entered:** 51 **Dropout:** NR **Data Loss:** • Home: 3 from unuse of portable system; 3 from faulty memory chip in portable • In-lab: 3 to technician procedure error; 4 due to error made when data was being uploaded **Analyzed:** 45 completed (Note: 40 of 44 in-lab recordings were split night studies)

Evidence Table 4. Reichert et al., 2003: Study Design, cont'd

Demographics	PSG Equipment	PSG Interpretation	Portable Interpretation
Males: 38 **Mean BMI:** 30 Range: 22-47 SD: ± 1.0 **Mean Age:** 52 Range: 30-83 SD: ±2.1 **% Non-White:** NR **Prevalence of OSA by PSG:** AHI>15: 20/44 Ave RDI: NR SD: NR **Comorbidities:** NR	**Setting:** Sleep lab **Equipment & Methodology:** Grass polygraph and recorded by Sandman Diagnostics System **Channels:** 2 EEG, EOG, submental EMG, ECG, anteior tibialis EMG, diagphragmatic EMG, microphone (snoring sounds), end tidal CO2, nasal –oral airflow, abdominal and thoracic respiration, and oximetry **Flow:** Thermocouple **Effort:** Piezo sensors	**Scoring:** Manual **Blinding:** Technologist blinded **Criteria for Excluding Data:** NR **Qualifications:** Trained	**Scoring:** Automated **Blinding:** For in lab: Technologist blinded to portable signals during recording and scoring of data **Data Quality Categorized:** NR **Criteria for Excluding Data:** NR **Qualifications:** NR

Evidence Table 4. Reichert et al., 2003: Study Design, cont'd

Comparison of Results	Special Statistics
Definitions: • Apnea: cessation of airflow for 10 s or longer • Hypopnea: ≥ 50% reduction in airflow for 10s or longer accompanied by a ≥ 2% decrease in oxygen hemoglobin saturation • AHI >15: positive for OSA (Home study data is averaged across the three nights) • AHI PSG: based on total sleep time; for Portable: based on total recording time; for home study it is the average across all nights of home testing **AHI/RDI:** • Avg PSG: NR • Portable: NR For AHI ≥ 15 (overall agreement 0.932) • Sensitivity – In lab: 95% ± 5% ;Home: 91% ± 6% • Specificity – In lab: 91% ± 6% ; Home: 83% ± 8% • Kappa: 0.864 ± 0.076 SE • LR+ 5.35, LR- 0.11 (calculated by reviewer) • Positive predictive value – In lab: 91% ± 6% ; Home: 83% ± 8% • Negative predictive value – In lab: 96% ± 4% ; Home: 91% ± 6% For RDI>5: • Home vs PSG Lab: Kappa: 0.734 ± 0.101 SE	**Bland-Altman Plot:** The differences between the PSG and home study are mainly clustered around the 0 horizontal line, with a heavier concentration for those when the mean AHI is less that 20. There were larger differences when the AHI was > 20 (As interpreted by the reviewer, though the mean difference appeared to be small, the limits of agreement estimated by ± 2 standard deviations (SD) were ± 60 events per hour) **ROC:** NR

Evidence Table 4. Reichert et al., 2003: Study Design, cont'd

Data Loss/ Variability/Cost Benefit	Quality Indicators	Author's Conclusions/ Limitations
Data Loss: 3 subjects used home portable less than three nights; 3 home studies had a faculty memory chip; 3 in-lab studies had technician procedure error and 4 in-lab studies had an error when data was being uploaded. **Night to Night Variability:** NR **Cost Benefit Analyses:** NR	**Evidence Level:** • Subjects selected consecutively or randomly (no design influence): N • Both tests scored blindly: Y • Reference standard done on all subjects: Y **Quality Level:** • Prospective study: Y • Portable done outside the lab: Y • Random order of allocation of subjects to PSG or Portable first: L-NA; H-N • Low data loss $\leq$ 10% of flow/SaO_2/EEG as applicable: N • High % completed ($\geq$ 90% of those entered): Y • PSG methodology fully described: Y • Portable monitor methodology fully described: N • Portable scoring fully described: N **Evidence Level Score:** II **Quality Level Score:** B (lab) D (home) **Grade:** Fair (lab) Poor (home)	**Conclusions:** • In populations suspected of having OSA, the portable demonstrated acceptable sensitivity and specificity both in the lab and self-administered in the home when compared to PSG • The portable was capable of accurately determining negative cases in both the lab and the home • There were 4 false positives in the home, but if the cutoff was changed to 18, they would all be eliminated • Two subjects had false negatives, but one of the subjects had all hypopneas, and the effect could be due to night-to-night variability • The in-home test corrobated the diagnoses and gave an indication of severity **Limitations:** • In lab night and home nights were not the same • Patient population did not allow testing whether the sensitivity would remain high in subjects with a low probability of OSA • Portable does not have a body position sensor, therefore position-related apnea can not be determined • The portable does not record EEG therefore sleep onset is not definitely known and sleep staging is not possible

Evidence Table 5. Ayas et al., 2003: Study Design

Study	Portable Device/ Implementation	Sample Selection	Study Subjects n (% of Entered)
Reference: Ayas NT, Pittman S, MacDonald M et al. Assessment of wrist-worn device in the detection of obstructive sleep apnea *Sleep Medicine*, 2003[13] **Study Location/Sponsor:** Brigham and Women's Hospital, Boston, MA **Type:** 4	**Device:** Watch_PAT100 (WPAT100) **Equipment & Methodology:** Uses a finger mounted optic/pneumatic sensor that eliminates venous pulsations and continuously measures the pulse volume of the digit **Channels:** 4: Peripheral arterial tone (PAT), oxygen saturation, heart rate, actigraphy **Flow:** NR **Effort:** NR **Portable Setting:** Sleep lab **Mode:** NR **Hook-up:** NR **Timing:** Simultaneous	**Site:** Sleep Center **Inclusion:** Adults suspected of having sleep apnea and persons without suspected sleep apnea recruited from advertisements **Exclusion:** Individuals using home oxygen, medically unstable, or using medications that block the alpha receptor **Clinical Pretest Probability Distribution of Suspected Severity of OSA:** NR	**Eligible:** 30 (25 with suspected sleep apnea, 5 healthy volunteers responding to a flyer) **Excluded:** NR **Entered:** NR **Dropout:** NR **Data Loss:** NR **Analyzed:** 30

Evidence Table 5. Ayas et al., 2003: Study Design, cont'd

Demographics	PSG Equipment	PSG Interpretation	Portable Interpretation
Males: 19 **Mean BMI:** 31.0 Range: NR SD: ± 7.6 kg/m2 **Mean Age:** 47 Range: 24.3-66.6 SD: ± 14.8 **% Non-White:** NR **Prevalence of OSA by PSG:** RDI>15: NR Ave RDI: NR SD: NR **Comorbidities:** NR	**Setting:** In-lab **Equipment & Methodology:** ALICE3 digital PSG system **Channels:** EOG, ECG, submental and tibial EMG, EEG (C2-A1, C3-A2, O1-A2, O2-A1), arterial oxygen saturation, snoring intensity, nasal-oral airflow, chest and abdominal motion **Flow:** Thermistor and nasal pressure **Effort:** NR	**Scoring:** Manual by single technologist **Blinding:** To portable data **Criteria for Excluding Data:** NR **Qualifications:** NR	**Scoring:** Automated **Blinding:** NR **Data Quality Categorized:** Respiratory events were detected during sleep, per actigraphy, using a combination of PAT signal attenuation, desaturation on pulse oximetry, and changes in heart rate. Criteria: ≥30% PAT amplitude reduction together with a pulse rate acceleration of 10% or ≥ 30% PAT amplitude reduction together with a 3% oxyhemoglobin desaturation, or a 4% oxyhemoglobin desaturation **Criteria for Excluding Data:** NR **Qualifications:** NR

Evidence Table 5. Ayas et al., 2003: Study Design, cont'd

Comparison of Results	Special Statistics
Definitions: • Apnea: scored if airflow was absent for 10 s • Hypopnea: if airflow was reduced by 50% or a lesser extent in association with a desaturation of 3% or an arousal • AHI: agreement if both were greater than 40 events per hour or if AHI was < per h on PSG and portable was within 10 events per h of AHI **AHI/RDI:** • Avg PSG: 23 + 23.9 events per hour, range: 1-94 • Portable: 23 ± 15.9 events per hour, range: 6-69 • Corr: Pearson's coefficient= 0.87, P <0.001 **Other Indices:** PLM index: mean=6 **Sleep Efficiency:** NR	**Bland-Altman Plot:** At lower levels of AHI, the portable tended to overestimate disease severity, while at higher levels of AHI, the portable underestimated severity **ROC:** • Set thresholds at 10,15,20 and 30 events per hour • For all, the AUC was greater than 0.86 • Optimal combinations of sensitivity/specificity: ♦ >10: 82.6/71.4 ♦ >15: 93.3/73.3 ♦ >20: 90.9/84.2 ♦ >30: 83.3/91.7

Evidence Table 5. Ayas et al., 2003: Study Design, cont'd

Data Loss/ Variability/Cost Benefit	Quality Indicators	Author's Conclusions/ Limitations
Data Loss: NR **Night to Night Variability:** NA **Cost Benefit Analyses:** NR	**Evidence Level:** • Subjects selected consecutively or randomly (no design influence): N • Both tests scored blindly: Y • Reference standard done on all subjects: Y **Quality Level:** • Prospective study: Y • Portable done outside the lab: N • Random order of allocation of subjects to PSG or Portable first: NA • Low data loss ≤ 10% of flow/SaO_2/EEG as applicable : N • High % completed (≥ 90% of those entered): Y • PSG methodology fully described: N • Portable monitor methodology fully described: Y • Portable scoring fully described: N **Evidence Level Score:** II **Quality Level Score:** D **Grading:** Poor	**Conclusions:** • Portable tended to overestimate events when PSG AHI was in the lower range (a PSG AHI of 0 would result in a portable of 5.6; 5, 9.5; 10, 13.3) but this may due to portable detecting events that were missed by standard scoring criteria or sleep fragmentation (arousals) • As PSG AHI increased, the portable tended to underestimate AHI (30, 28.7; 60, 51.8) because portable had more difficulty detecting each individual event when such multiple events occurred over a brief time period **Limitations:** • Did not test Watch_PAT in the home, so performance under these conditions is not known • Can't identify patients with other disorders of sleep that lead to substantial sleep fragmentation and arousals • Device does not measure airflow so can not differentiate hypopneas from apneas • Small sample, one sleep center • Did not perform an event-by-event analysis of portable vs PSG AHI

Evidence Table 6. Bar et al., 2003: Study Design

Study	Portable Device/ Implementation	Sample Selection	Study Subjects n (% of Entered)
Reference: Bar A, Pillar G, Dvir I. Evaluation of a portable device based on peripheral arterial tone for unattended sleep studies. *Chest*, 2003[14] **Study Location/ Sponsor:** Haifa, Israel/ Itamar Medical Ltd. **Type:** 4	**Device:** Watch PAT100(WP100) **Equipment & Methodology:** Portable device based on peripheral arterial tone (PAT) signal and designed for unattended home sleep studies; battery powered console unit mounted just above the wrist and two finger mounted probes **Channels:** 4: actigraphy, peripheral arterial tone (PAT), oximetry, and pulse rate (derived from PAT signal) **Flow:** NA **Effort:** NA **Portable Setting:** • In-lab • Substudy in Home **Mode:** • In-lab: NR • Substudy in Home: unattended **Hook-up:** • In-lab: NR • Substudy in Home: patient **Timing:** • In-lab: simultaneous • Home: two additional nights	**Site:** Sleep Lab **Inclusion:** Referred adults with suspected OSAS; home study only: within 30 mile range of sleep laboratory **Exclusion:** Permanent pacemaker, nonsinus cardiac arrythmias, peripheral vasculopathy or neuropathy, severe lung disease, status postbilateral cervical or thoracic sympathectomy, finger deformity that precludes adequate sensor application, use of adrenergic receptor blockers (24h washout period required) and alcohol or drug abuse during the last 3 years **Clinical Pretest Probability Distribution of Suspected Severity of OSA:** NR	**Eligible:** 69 consecutive subjects referred to clinic and 33 healthy volunteers **Excluded:** NR **Entered:** • In-lab: 102; • Substudy in Home: 14 (subset of in-lab) **Dropout:** 0 **Data Loss:** • In-lab: 3 studies: 2 PSG studies had synchronization failure; 1 portable had valid sleep time <1.5 h • Substudy in Home: 3 of 28 portable studies due to technical failure but were repeated successfully **Analyzed:** • In-lab: 99 • Substudy in Home: 28

Evidence Table 6. Bar et al., 2003: Study Design, cont'd

Demographics	PSG Equipment	PSG Interpretation	Portable Interpretation
Males: 78 **Mean BMI:** 26.8 Range: NR SD: ±5.5 **Mean Age:** 41.4 Range: NR SD: ±15.2 **% Non-White:** NR **Prevalence of OSA by PSG:** 67.6% (69/102) plus 33 normal volunteers **Comorbidities:** 20% hypertension 4% coronary heart disease	**Setting:** Sleep Lab **Equipment & Methodology:** Embla **Channels:** 2 EEG (C3-A2, O2-A1), right and left EOG, chin EMG, arterial oxygen saturation, nasal-oral airflow, ECG, chest and abdominal wall motion, bilateral anterior tibialis EMG, and body position **Flow:** Thermisters and nasal pressure **Effort:** Piezo electrodes	**Scoring:** Manual **Blinding:** To portable data and results **Criteria for Excluding Data:** AASM criteria • PSG sleep time<1.5h • Technical failure of synchronizing the PSG to the WP100 • Poor quality of the PSG recording **Qualifications:** NR	**Scoring:** Automated (proprietary algorithm) **Blinding:** To PSG data **Data Quality Categorized:** NR **Criteria for Excluding Data:** • Technical failure of synchronizing the PSG to the WP100 • WP100-related rejection (valid sleep time < 1.5 h) **Qualifications:** NR

Evidence Table 6. Bar et al., 2003: Study Design, cont'd

Comparison of Results	Special Statistics
Definitions: • Apnea/hypopnea event: airflow amplitude but with the presence of arousal of oxygen desaturation of at least 3% • PSG RDI: # of apnea/hypopnea per hour of sleep • OSA diagnosis: PSG-RDI>10 **AHI/RDI:** • In-lab portable versus PSG: ♦ Corr: r=0.88, P<0.0001, n=99 • Portable in-lab versus portable in-home: ♦ Mean of two nights of home study: Corr: r=0.89, P<0.001, n=14 ♦ Each night of home study: Corr: r=0.94, P<0.001, n=14 • For RDI ≥ 20 (as calculated by reviewer): ♦ PSG: 9 ♦ Portable: 7/9 ♦ Sensitivity: 78% ♦ Specificity: 60% (two false-positives) (1 false positive had PSG RDI of 19 and if considered correctly classified, then sensitivity would rise to 80% and specificity to 75%) ♦ LR+: 3.2; LR-: 0.27	**Bland-Altman Plot:** There was a slight tendency for the portable to underscore events in the mild range of OSAS and to overscore events in the severe range **ROC:** • With the RDI threshold set at 10, the diagnostic threshold, the area under the curve is 0.82 (P=0.0001) showing the potentially high sensitivity and specificity in diagnosing OSAS • With the RDI threshold set at 20, when CPAP therapy is indicated, the area under the curve is 0.87 (P=0.0001) showing the potentially high sensitivity and specificity in diagnosing OSAS of a severity requiring treatment

Evidence Table 6. Bar et al., 2003: Study Design, cont'd

Data Loss/ Variability/Cost Benefit	Quality Indicators	Author's Conclusions/ Limitations
Data Loss: In-lab: 3 of 102 Home: 3 of 28 **Night to Night Variability:** NA **Cost Benefit Analyses:** NR	**Evidence Level:** • Subjects selected consecutively or randomly (no design influence): Y • Both tests scored blindly: Y • Reference standard done on all subjects: ♦ In Home Substudy 3 tests were repeated: Y **Quality Level:** • Prospective study: Y • Portable done outside the lab: L-N; H-Y • Random order of allocation of subjects to PSG or ♦ Lab: NA ♦ Home: N • Low data loss ≤ 10% of flow/SaO_2/ EEG as applicable: L-Y; H-N • High % completed (≥ 90% of those entered): Y • PSG methodology fully described: Y • Portable monitor methodology fully described: Y • Portable scoring fully described: N **Evidence Level Score:** II **Quality Level Score:** C (lab) C (home) **Grade:** Fair (lab) Fair (home)	**Conclusions:** • The in-lab portable comparison with PSG results were well correlated with good efficacy for both OSAS diagnosis and therapy indication • The in-lab portable results were highly reproducible in the home sleep studies, for laboratory versus home and between the two home sleep studies • Device reliability is demonstrated by low rejection rates **Limitations:** • Study population consisted of patients with snoring/sleep apnea syndrome and healthy volunteers • Did not evaluate the device for upper airway resistance syndrome

Evidence Table 7. Golpe et al., 2002: Study Design

Study	Portable Device/ Implementation	Sample Selection	Study Subjects n (% of Entered)
Reference: Golpe R, Jimenez A, Carpizo R. Assessment of sleep apnea/hypopnea syndrome. *Chest,* 2002. 122(4); 1156-1161[11] **Study Location/ Sponsor** Spain/NR **Type:** 4	**Device:** Apnoescreen-I **Equipment & Methodology:** NR **Channels:** Body position, wrist actimetry, pulse rate, arterial oxygen saturation **Flow:** Oronasal airflow with thermister **Effort:** NA **Portable Setting:** Home **Mode:** Unattended **Hook-up:** 50/50 split: • G1: by Technician at home • G2: by patients after 15-20 minute training period at the hospital and written instructions **Timing:** portable done first (PSG done within 30 days)	**Site:** Sleep lab **Inclusion:** At least 2 of the following: loud snoring, observed apneas, daytime drowsiness; and judged by author to be require sleep study **Exclusion:** Physical or mental impairment that precluded use of equipment **Clinical Pretest Probability Distribution of Suspected Severity of OSA:** NR	**Eligible:** 59 invited **Excluded:** 0 **Entered:** • 55, randomly assigned ♦ G1: 28 ♦ G2: 27 **Dropout:** NR **Data Loss:** • 11 (20%) home studies produced no interpretable data ♦ G1: 2 (poor signals in flow channel) ♦ G2: 9 (3 where patient failed to turn on equipment and 6 due to poor signals or artifacts in the flow channel) **Analyzed:** 44

Evidence Table 7. Golpe et al., 2002: Study Design, cont'd

Demographics	PSG Equipment	PSG Interpretation	Portable Interpretation
Males: 96% **Mean BMI:** 30.3 Range: NR SD: 4.6 **Mean Age:** 52.7 Range: NR SD: 13.3 **% Non-White:** NR **Prevalence of OSA by PSG:** RDI>10: 23/44 Ave RDI: NR SD: NR **Comorbidities:** NR	**Setting:** Sleep lab **Equipment & Methodology:** Medelec **Channels:** 14: EEG, Chin electromyogram, electro-oculogram, ECG, tibial electromyograms, oxygen saturation with a finger sensor, body position, snoring, oronasal, and thoracoabdominal movement **Flow:** Oronasal by thermister **Effort:** Piezo-electric bands	**Scoring:** Manual **Blinding:** Investigator blinded to home study results and other investigators results Note: Investigators had intake interview results available to them **Criteria for Excluding Data:** NR **Qualifications:** NR, but one of the authors, an MD was an investigator	**Scoring:** Automated and manual (through counting episodes on the video screen graphic display-which does not measure with accuracy the Type of desaturation) **Blinding:** Investigator blinded to in-hospital study results and other investigators results Note: Investigators had intake interview results available to them **Data Quality Categorized:** NR **Criteria for Excluding Data:** NR **Qualifications:** NR, but one of the authors, an MD was an investigator

Evidence Table 7. Golpe et al., 2002: Study Design, cont'd

Comparison of Results	Special Statistics
Definitions: • Apnea: complete cessation of airflow ≥ 10 s • Hypopnea: discernable reduction in respiratory airflow ≥ 10s and SaO_2 decreased ≥ 4% oxygen saturation and/or arousal (thermisters only allow qualitative) • PSG AHI: average # of episodes of apnea and hypopnea per hour of sleep • Portable: average # of episodes of apnea and hypopnea divided by the registry time and sleep time in hours • Manual respiratory index per hour of registry time (mRDI-r) ≥ 10 = SAHS PSG SAHS: n=23 (including "doubtful" home studies) • Home study: 18/23 in agreement • False Negatives: 1/23 PSG No SAHS: n=21 (including "doubtful" home studies) • Home study: 15/21 in agreement • False Positives: 3/21 PSG SAHS: n=19 (without "doubtful" home studies) • Home study: 18/19 in agreement • False Negatives: 1/19 PSG No SAHS: n=18 (without "doubtful" home studies) • Home study: 15/18 in agreement • False Positives: 3/18	**Bland-Altman Plot:** Horizontal lines were drawn at the mean difference: -4.2±1.96 the SD of the difference. The plot shows a large spread for the Type of agreement and the text notes that there were 5 cases were outside the limit of agreement **ROC:** There were no differences between the areas under the curves when the respiratory disturbances indexes were referred either to registry time or to the estimated sleep time. The manual respiratory disturbance (either referred to the registry time or to the estimated sleep time) seemed the most useful indexes for the diagnosis of SAHS, as they had the larger area under the ROC curve, but the values obtained for the indexes were too similar to assess if the differences was significant. No other curve details presented AUC: 0.89 for manual scoring and 0.86 for automated scoring. (Reviewer visual interpretation of ROC found that best sensitivity was 90% and best specificity was 80% which calculated to a LR+ of 4.5 and an LR- of 0.125)

Evidence Table 7. Golpe et al., 2002: Study Design, cont'd

Data Loss/ Variability/Cost Benefit	Quality Indicators	Author's Conclusions/ Limitations
Data Loss: G1: 7% G2: 33% **Night to Night Variability:** NR, but mentioned **Cost Benefit Analyses:** Home studies with a technician setup of the equipment were less expensive (because of the high percentage of faulty studies with patient own setup of sleep recording devices)	**Evidence Level:** • Subjects selected consecutively or randomly (no design influence): Y • Both tests scored blindly: Y • Reference standard done on all subjects: Y **Quality Level:** • Prospective study: Y • Portable done outside the lab: Y • Random order of allocation of subjects to PSG or Portable first: N • Low data loss $\leq$ 10% of flow/SaO_2/EEG as applicable : N • High % completed ($\geq$ 90% of those entered): Y • PSG methodology fully described: Y • Portable monitor methodology fully described: Y • Portable scoring fully described: N **Evidence Level Score:** I **Quality Level Score:** C **Grade:** Fair	**Conclusions:** • Home studies are a viable form of diagnosing SAHS because they are more economic even with relatively high %s of faulty or inconclusive recording that lead to repeat studies. Technician intervention still is less costly than in-hospital PSG. • Agreement on the clinical decision on whether to treat the patient with CPAP based on the home study result was 77% (34 of 44 cases) **Limitations:** • Did not analyze the long-term results of CPAP therapies • Social Type of population should be taken into account • Home study overestimated the severity of SAHS. There were 3 false negatives and 3 false positives, though authors suggest the latter 3 could be false-negatives from the PSG or night-to-night variability. • Relative complexity of home study systems are useful for rural areas but these patients may not be as familiar with electronic devices

Evidence Table 8. Gurubhagavatula et al., 2004: Study Design

Reference	Portable Device/ Implementation	Sample Selection	Study Subjects n (% of Entered)
Gurubhagavatula I, Maislin G, Nkwuo JE et al. Occupational screening for obstructive sleep apnea in commercial drivers. *American Journal of Respiratory Care,* 2004[20] **Study Location/ Sponsor:** Philadelphia, PA, US/ Trucking Research Institute of the American Trucking Institute, funded by the Federal Motor Carriers Safety Administration; and Nellcor, Inc., and Ohmeda, Inc provided partial support **Type:** 4	**Device:** NR **Equipment & Methodology:** Nocturnal Pulse Oximetry **Channels:** 1: oximetry **Flow:** NR **Effort:** NR **Portable Setting:** In-lab **Mode:** NR **Hook-up:** NR **Timing:** Simultaneous	**Site:** Philadelphia, PA **Inclusion:** People with commercial drivers licenses within a 50 mile radius of Philadelphia, PA (random selection) **Exclusion:** NR **Clinical Pretest Probability Distribution of Suspected Severity of OSA:** 0.49(0.21)	**Eligible:** 4,286 questionnaires mailed **Excluded:** Non-responders to the questionnaire **Entered:** 1,329 enrolled, stratified into high and low risk for apnea. Performed oximetry and PSG in 44.8% (247/551) in high risk group and 20.4% (159/778) in low risk group **Dropout:** NR **Data Loss:** NR **Analyzed:** 1,329

Evidence Table 8. Gurubhagavatula et al., 2004: Study Design, cont'd

Demographics	PSG Equipment	PSG Interpretation	Portable Interpretation
Males: 93.5% **Mean BMI:** 28.4 Range: SD: ± 4.85 Approximately half of the sample was obese (≥30 kg/m^2) **Mean Age:** 44.4 Range: SD: ± 11.2 **% Non-White:** 15% **Prevalence of OSA by PSG:** (weighted sample) AHI ≥ 5: 28.1% AHI ≥ 30: 4.7% Ave RDI: NR SD: NR **Comorbidities:** Current smoking: 30.9 % (±2.5) Any smoking: 61% (±2.6)	**Setting:** In-lab **Equipment & Methodology:** NR **Channels:** NR **Flow:** NR **Effort:** NR	**Scoring:** NR **Blinding:** To other results **Criteria for Excluding Data:** NR **Qualifications:** NR	**Scoring:** NR **Blinding:** To other results **Data Quality Categorized:** NR **Criteria for Excluding Data:** NR **Qualifications:** NR

Evidence Table 8. Gurubhagavatula et al., 2004: Study Design, cont'd

Comparison of Results	Special Statistics
Definitions: Any apnea: AHI≥5 (upper bound of 0.9, lower bound 0.2 and desaturation threshold of 5 events per hour); Severe apnea: AHI≥30 (upper bound of 0.9, lower bound 0.3, desaturation threshold of 10 events per hour) **AHI/RDI:** Oximetry versus PSG: • For AHI ≥ 15 ♦ Sensitivity – .89 (0.738-1.000) ♦ Specificity – .95 (0.897-0.951)	**Bland-Altman Plot:** NR **ROC:** • Multivariable prediction plus oximetry with 0.9,0.3 and 10± cut points had an area under the curve of 0.937 (0.936-0.939) and Likelihood Ratio of 0.100 (0.035-0.323) • Oximetry alone prediction with a 14.9 cut point had an area under the curve of 0.971 (0.945-0.992) and a Likelihood Ratio of 1.22 (0.000-0.283)

Evidence Table 8. Gurubhagavatula et al., 2004: Study Design, cont'd

Data Loss/ Variability/Cost Benefit	Quality Indicators	Author's Conclusions/ Limitations
Data Loss: NR **Night to Night Variability:** NR **Cost Benefit Analyses:** Only in regard to motor vehicle crashes	**Evidence Level:** • Subjects selected consecutively or randomly (no design influence): Y • Both tests scored blindly: N • Reference standard done on all subjects: Y **Quality Level:** • Prospective study: Y • Portable done outside the lab: N • Random order of allocation of subjects to PSG or Portable first: NA • Low data loss $\leq$ 10% of flow/SaO_2/EEG as applicable: N • High % completed ($\geq$ 90% of those entered): N • PSG methodology fully described: N • Portable monitor methodology fully described: N • Portable scoring fully described: N **Evidence Level Score:** II **Quality Level Score:** D **Grade:** Poor	**Conclusions:** • The oximetry enhanced (symptoms + BMI) approach had a sensitivity of 89% and specificity of 95% and had better discriminatory power (negative likelihood , 0.29) • Neither oximetry strategy excluded drivers with AHI≥5 with acceptable sensitivity • For drivers with AHI≥30, the enhanced oximetry had a 91% sensitivity and specificity and yielded a 0.10 likelihood ratio; PSG was not necessary in 86%, the false-positive rate was 8.9% and the false-negative rate was 0.5% and the negative predictive value of 99%; oximetry did not improve the negative likelihood ratio, thus to exclude severe apnea, the enhanced oximetry approach is optimal • For any apnea, oximetry offered minimal additional predictive advantage with a negative likelihood ratio of 0.29 **Limitations:** • Validated strategies in the same cohort in which it was derived which may artificially inflate its predictive value • Oximetry was not conducted in the home

Evidence Table 9. Liesching et al., 2004: Study Design

Study	Portable Device/ Implementation	Sample Selection	Study Subjects n (% of Entered)
Reference: Liesching TN, Carlisle C, Marta A et al. Evaluation of the accuracy of SNAP technology sleep sononography in detecting obstructive sleep apnea in adults compared to standard polysomnography. *Chest,* 2004[19] **Study Location/ Sponsor:** Rhode Island, USA/NR **Type:** 4	**Device:** SNAP technology **Equipment & Methodology:** Home system consists of a microphone cannula device that is placed on the subject's upper lip during sleep **Channels:** Oronasal respiratory sound and airflow **Flow:** Monitors positioned over the oral and nasal apertures **Effort:** NR **Portable Setting:** Home **Mode:** Unattended **Hook-up:** Patient **Timing:** Mean followup time between the two studies was 5 months (range 2 to 10 months)	**Site:** Sleep Clinic **Inclusion:** Referred to the sleep center for PSG; prior receipt of SNAP technology testing in past 2 years based on referral of primary care physician for suspected sleep apnea (per retrospective medical record review) **Exclusion:** Not having had SNAP technology testing in past 2 years; patients who had SNAP testing but refused subsequent PSG **Clinical Pretest Probability Distribution of Suspected Severity of OSA:** NR	**Eligible:** 39 **Excluded:** 3 refused further diagnostic evaluation or workup **Entered:** 36 **Dropout:** 4 were referred for split-night studies at request of physician **Data Loss:** 1 SNAP test did not provide an AHI value **Analyzed:** 31

Evidence Table 9. Liesching et al., 2004, cont'd

Demographics	PSG Equipment	PSG Interpretation	Portable Interpretation
Males: 14 of 31 **Mean BMI:** 31.6 kg/m^2 Range: 24-44 SD: NR **Mean Age:** 50.3 Range: 29 - 77 SD: NR **% Non-White:** NR **Prevalence of OSA by PSG:** AHI>5: 45% Ave RDI: NR SD: NR **Comorbidities:** NR	**Setting:** Sleep Disorders Center **Equipment & Methodology:** NR **Channels:** EEG, EOG, submental EMG, chest and abdominal impedence, intercostals EMG, airflow, SaO_2 through pulse oximetry, ECG **Flow:** Thermisters and nasal pressure transducer **Effort:** Plethysmography	**Scoring:** Manual **Blinding:** To results of SNAP studies **Criteria for Excluding Data:** NR **Qualifications:** Physician trained in sleep medicine	**Scoring:** NR **Blinding:** NR **Data Quality Categorized:** NR **Criteria for Excluding Data:** NR **Qualifications:** NR

Evidence Table 9. Liesching et al., 2004

Comparison of Results	Special Statistics
Definitions: • PSG: apnea/hypopnea event: clear decrease(>50%) from baseline in the amplitutde of the nasal pressure transducer signal during sleep, or by a clear amplitude reduction in the nasal pressure transducer signal during sleep that does not reach the >50% criterion but is associated with either an oxygen desaturation of >3% or an arousal, in either case the event must last ≥10 s • Apnea in portable: cessation of sound for > 10 s • Hypopnea in portable: sound amplitude is reduced to <25% of the baseline (i.e., quiet respiration) amplitude for at least 10 s. Baseline sound can not be secondary to snoring; or reduction in sound is part of a cyclical change in amplitude; or respiratory rate is not significantly reduced during event • AHI: total number of apneas and hypopneas per hour of sleep (PSG) per hour of the study (Portable) • OSA: PSG ≥ 5 events per hour; mild: 5 to 15; moderate: 15 to 30; severe: >30 **AHI/RDI:** • Mean PSG: 21.2±21.5; Median: 11.5 per hour • Portable: 19.8±14.8; Median: 15.8 per hour ♦ P=0.42 • Corr: Kappa: 0.23 (p=0.008) • Portable accurately assessed true severity in 38.7% (12/31); 54.8% (17/31) accurately predicted AHI within 10 events per hour; 45.2% (14/31) the difference was greater than 10 events per hour • Of 8 normal PSG results, the portable rated these studies as 6 mild, 1 moderate, and 1 severe, or a specificity of 0%. Using only the moderate and severe OSA studies from the portable, the specificity increases to 75%. • 1 false negative by the portable was classified as severe by the PSG. • Including all false positive cases in the specificity, the LR+ is 0.91 (calculated by reviewer) • For the PSG, normal patients misclassified as moderate or severe by the portable, the LR+ is 3.64 and the LR- is 12 (calculated by reviewer) • Portable correctly classified 21 of 23 positive PSG OSA cases, or 91% sensitivity as calculated by the reviewer	**Bland-Altman Plot:** NR **ROC:** NR

Evidence Table 9. Liesching et al., 2004, cont'd

Data Loss/ Variability/Cost Benefit	Quality Indicators	Author's Conclusions/ Limitations
Data Loss: 1 of 32 (3%) **Night to Night Variability:** NR **Cost Benefit Analyses:** NR	**Evidence Type:** • Subjects selected consecutively or randomly (no design influence): N • Both tests scored blindly: Y • Reference standard done on all subjects: Y **Quality Type:** • Prospective study: N • Portable done outside the lab: Y • Random order of allocation of subjects to PSG or Portable first: N • Low data loss ≤ 10% of flow/SaO_2/EEG as applicable : Y • High % completed (≥ 90% of those entered): N • PSG methodology fully described: Y • Portable monitor methodology fully described: Y • Portable scoring fully described: N **Evidence Type Score:** II **Quality Type Score:** D **Grade:** Poor	**Conclusions:** • SNAP severity scores overestimated in 41.9% of patients, 25.8% of which had normal PSGs, two patients had moderate to severe AHI results from SNAP • SNAP study underestimated severity in 19.4% (6 of 31) by at least 10 events per hour **Limitations:** • Sensitivity and specificity could not be calculated because study population did not include those patients that were tested by SNAP but not referred to the sleep clinic • SNAP device does not measure sleep position • Studies were not done simultaneously and in two different settings and two different methods of hook-up and monitoring

Evidence Table 10. Pillar et al., 2003: Study Design

Study	Portable Device/ Implementation	Sample Selection	Study Subjects n (% of Entered)
Reference: Pillar G, Bar A, Betito M et al. An automatic ambulatory device for detection of AASM defined arousals from sleep: the WP100. *Sleep Medicine*, 2003[17] **Study Location/ Sponsor:** Technion Sleep Disorders Center, Israel/ NR **Type:** 4 Note: Authors Bar, Betito, Schnall, Sheffy are employees of Itamar Medical and Authors Pillar and Lavie are consultants for Itamar Medical Ltd., makers of the device	**Device:** Watch_PAT100 (WP100) **Equipment & Methodology:** Battery powered consol unit mounted just above the wrist. Actigraph and oximetry embedded into device **Channels:** 4: actigraphy, peripheral arterial tone (PAT), oximetry, and pulse rate (derived from PAT signal) **Flow:** NR **Effort:** NA **Portable Setting:** Sleep lab **Mode:** NR **Hook-up:** NR **Timing:** Simultaneous	**Site:** Sleep Clinic **Inclusion:** Referred for presumed OSAS in 61 and 7 healthy volunteers **Exclusion:** Permanent pacemaker, non-sinus cardiac arrhythmias, peripheral vasculopathy or neuropathy, severe lung disease, S/P bilateral cervical or thoracic sympathectomy, finger deformity that precludes adequate sensor application, use of alpha-adrenergic receptor blockers (24h washout period required), and alcohol or drug abuse during the last 3 years **Clinical Pretest Probability Distribution of Suspected Severity of OSA:** NR	**Eligible:** 68: 61 consecutive patients referred to clinic and 7 healthy volunteers responding to a flyer **Excluded:** NR **Entered:** NR **Dropout:** NR **Data Loss:** NR **Analyzed:** 68

Evidence Table 10. Pillar et al., 2003: Study Design, cont'd

Demographics	PSG Equipment	PSG Interpretation	Portable Interpretation
Males: 54 **Mean BMI:** 28 Range: NR SD: ± 6kg/m2 **Mean Age:** 46 Range: NR SD: ± 14 years **% Non-White:** NR **Prevalence of OSA by PSG:** RDI>15: NR Ave RDI: 34 SD: ± 26 events per hour **Comorbidities:** 20% hypertension 4% coronary artery disease	**Setting:** Sleep lab **Equipment & Methodology:** Computerized PSG **Channels:** 2 EEG (C3-A2, O2-A1), EOG, submental EMG, SaO_2, nasal-oral airway, EKG, chest and abdominal wall motion, bilateral anterior tibialis EMG, and body position **Flow:** Thermisters and nasal pressure **Effort:** Piezo electrodes	**Scoring:** Manual **Blinding:** To other data and results **Criteria for Excluding Data:** NR **Qualifications:** NR	**Scoring:** Automatic **Blinding:** To other data and results **Data Quality Categorized:** NR **Criteria for Excluding Data:** NR **Qualifications:** NR

Evidence Table 10. Pillar et al., 2003: Study Design, cont'd

Comparison of Results	Special Statistics
Definitions: • Arousal: EEG frequency shift of ≥ 3 s ≤15 s during non-REM sleep, during REM sleep an increase in EMG was required as well, and in both cases at least 10 s of sleep prior to and following the event was required • Arousal Index: dividing total # of arousals by # of hours of sleep • RDI: # of apneas plus hypopneas divided by the # of hours of sleep **AHI/RDI:** For RDI ≥ 20 • Sensitivity – 0.80 • Specificity – 0.79	**Bland-Altman Plot:** Across a wide range of arousal frequencies, there was good agreement between the portable arousal index and the PSG arousal index **Scatter Plot:** Very high and statistically significant correlation (r=0.87), P<0.0001) was found between the portable autonomic arousal index and the PSG arousal index **ROC:** Using threshold of 20 arousals per hour of sleep the area under the curve=0.87 • Sensitivity: 0.80 • Specificity: 0.79

Evidence Table 10. Pillar et al., 2003: Study Design, cont'd

Data Loss/ Variability/Cost Benefit	Quality Indicators	Author's Conclusions/ Limitations
Data Loss: NR **Night to Night Variability:** NA **Cost Benefit Analyses:** NR	**Evidence Level:** • Subjects selected consecutively or randomly (no design influence): Y • Both tests scored blindly: N • Reference standard done on all subjects: Y **Quality Score:** • Prospective study: Y • Portable done outside the lab: N • Random order of allocation of subjects to PSG or Portable first: NA • Low data loss $\leq$ 10% of flow/SaO_2/EEG as applicable: N • High % completed ($\geq$ 90% of those entered): Y • PSG methodology fully described: Y • Portable monitor methodology fully described: Y • Portable scoring fully described: N **Evidence Level Score:** IV **Quality Level Score:** C **Grade:** Poor	**Conclusions:** • Automatic analysis of PAT signal derived from the WPAT100 can accurately identify arousals from sleep **Limitations:** • Took place in a sleep lab • Aimed at detecting sleep fragmentation but chose to compare autonomic arousals to the commonly detected EEG arousals, but PAT may be sensitive to shorter arousals (<3 s) • Population was made up of patients with snoring/sleep apnea syndrome and a few healthy volunteers

Evidence Table 11. Shochat et al., 2002: Study Design

Reference	Portable Device/ Implementation	Sample Selection	Study Subjects n (% of Entered)
Reference: Shochat T, Hadas N, Kerkhofs M et al. The SleepStrip: an apnoea screener for the early detection of sleep apnoea syndrome. *European Respiration Journal,* 2002[12] **Study Location/ Sponsor:** Israel, Germany, Belgium/ NR **Type:** 4 Note: P. Lavie is the scientific advisor to the company manufacturing the SleepStrip; N. Hadas is the company's R&D manager; T. Shochat is the Clinical Trials Director of the company; both P. Lavie and N. Hadas are both shareholders in the company	**Device:** SleepStrip™ **Equipment & Methodology:** Disposable sleep apnoea screener designed to aid sleep specialists or other physicians in the initial assessment of patients with suspected SAS; it is small, light weight and worn underneath the nose and above the upper lip **Channels:** Oral and nasal flow sensors **Flow:** Thermisters **Effort:** NA **Portable Setting:** In-lab **Mode:** NR **Hook-up:** NR **Timing:** Concomitantly	**Site:** Assume sleep clinics in three locations: Israel, Belgium and Germany **Inclusion:** Patients referred with suspicion of sleep apnea **Exclusion:** NR **Clinical Pretest Probability Distribution of Suspected Severity of OSA:** NR	**Eligible:** NR **Excluded:** NR **Entered:** 402 (303 from Israel, 50 from Belgium, 49 from Germany) **Dropout:** NR **Data Loss:** Insufficient sleep (<5 hr): 22% (n=88); technical malfunction of portable: 8% (n=31); missing data: 1.5% (n=6); note: 11 cases had both insufficient sleep and technical malfunction **Analyzed:** 288 (72%) (205 from Israel, 39 from Belgium, 44 from Germany)

Evidence Table 11. Shochat et al., 2002: Study Design, cont'd

Demographics	PSG Equipment	PSG Interpretation	Portable Interpretation
Males: NR **Mean BMI:** NR Range: NR SD: NR **Mean Age:** NR Range: NR SD: NR **% Non-White:** NR **Prevalence of OSA by PSG:** RDI>15: NR Ave RDI: NR SD: NR **Comorbidities:** NR	**Setting:** Sleep Clinic **Equipment & Methodology:** NR **Channels:** Respiration, oronasal temperature to assess respiratory flow, oxygen saturation and percentage of sleep-time with oxygen saturation <90% **Flow:** NR **Effort:** Respiratory effort belts around the chest	**Scoring:** NR **Blinding:** NR **Criteria for Excluding Data:** NR **Qualifications:** Experienced scorers in each of the three settings	**Scoring:** Automatically derived scores **Blinding:** NA **Data Quality Categorized:** **Criteria for Excluding Data:** Device worn for less than 5 hours **Qualifications:** NA

Evidence Table 11. Shochat et al., 2002: Study Design, cont'd

Comparison of Results	Special Statistics
Definitions: • Apneas: cessation of airflow ≥10 s • Hypopneas: reduction in the amplitude for the respiratory signal of at least 50% for ≥ 10 s, followed by either a decrease in oxygen saturation or 4% or signs of physiological arousal • PSG AHI OSA categories: >10 = mild; >20= moderate; >40= severe **Apnea Index:** AHI/RDI: Avg PSG: NR • Portable: NR • Corr: Pearson r=0.73, P<0.001) For AHI/Sscore >10 • Sensitivity – 86% • Specificity – 57% For AHI/Sscore >20 • Sensitivity – 80% • Specificity – 70% For AHI/Sscore >40 • Sensitivity – 80% • Specificity – 86%	**Bland-Altman Plot:** There is a bias of -5.8 events (CI:-8 to -3.6 events) indicating a small overestimation of the severity of the apnoea by the portable **ROC:** Area under the curve ranged from 0.81 to 0.92 at the varying AHI thresholds: • >10: 0.81 (P<0.001) CI: 0.76-0.86 • >20: 0.84 (P<0.001) CI: 0.79-0.89 • >40: 0.92 (P<0.001) CI: 0.89-0.96

Evidence Table 11. Shochat et al., 2002: Study Design, cont'd

Data Loss/ Variability/Cost Benefit	Quality Indicators	Author's Conclusions/ Limitations
Data Loss: NR **Night to Night Variability:** NA **Cost Benefit Analyses:** NR	**Evidence Type:** • Subjects selected consecutively or randomly (no design influence): N • Both tests scored blindly: N • Reference standard done on all subjects: Y **Quality Type:** • Prospective study: Y • Portable done outside the lab: N • Random order of allocation of subjects to PSG or Portable first: NA • Low data loss ≤ 10% of flow/SaO_2/EEG as applicable: N • High % completed (≥ 90% of those entered): Y • PSG methodology fully described: N • Portable monitor methodology fully described: N • Portable scoring fully described: N **Evidence Type Score:** IV **Quality Type Score:** D **Grade:** Poor	**Conclusions:** • Though not intended as a substitute for PSG, the SleepStrip may provide initial screening information, which may be useful in both clinical and experimental settings • Patients with medical conditions which frequently coexist with SAS, such as ischaemic heart disease, COPD and ESRD may find this technology more appealing as it can be used in the comfort of their home **Limitations:** • Not intended to replace full PSG but is for large scale screening

Evidence Table 12. Zamarron et al., 2003: Study Design

Study	Portable Device/ Implementation	Sample Selection	Study Subjects n (% of Entered)
Reference: Zamarron C; Gude F; Barcala J et al. Utility of oxygen saturation and heart rate spectral analysis obtained from pulse oximetric recordings in the diagnosis of sleep apnea syndrome. *Chest,* 2003[18] **Study Location/ Sponsor:** Barcelona, Spain/ Fondo Investigacion Sanitaria; and Secretaria Xeral de Investigacion e Desenvolvemento **Type:** 4	**Device:** Criticare 504 Oximeter **Equipment & Methodology:** Finger probe pulse oximetry recording **Channels:** SaO_2 and heart rate **Flow:** NA **Effort:** NA **Portable Setting:** In-lab **Mode:** NR **Hook-up:** Assume technician **Timing:** Simultaneous	**Site:** Sleep Clinic **Inclusion:** Showing clinical symptoms of OSA and referred by a general practitioner **Exclusion:** NR **Clinical Pretest Probability Distribution of Suspected Severity of OSA:** NR	**Eligible:** 300 **Excluded:** NR **Entered:** 300 **Dropout:** NR **Data Loss:** NR **Analyzed:** 300

Evidence Table 12. Zamarron et al., 2003: Study Design, cont'd

Demographics	PSG Equipment	PSG Interpretation	Portable Interpretation
Males: 78% **Mean BMI:** 29.5 Range: NR SD: ±5.3 **Mean Age:** OSA: 58.4 ± 12.1; Non-OSA: 56.6 ±12.8 Range: 21 to 84 (total) SD: NR **% Non-White:** NR **Prevalence of OSA by PSG:** AHI>10: 56% Ave AHI: 40.2 SD: ±22.4 **Comorbidities:** • OSA group had significantly higher STOT, S 30-70 S, and P values than non-OSA group • COPD, Cardiovascular Disease	**Setting:** Sleep Clinic **Equipment & Methodology:** Ultrasom **Channels:** EEG, EOG, chin EMG, air flow, ECG, chest wall movement **Flow:** Three-part thermister **Effort:** NR	**Scoring:** NR, assume manual; analyzed in periods of 30 s and during stages 1,2,3,4 and rapid eye movement **Blinding:** Single; independent observer of the portable data **Criteria for Excluding Data:** If subject has < 3 h of total sleep, study was repeated **Qualifications:** NR	**Scoring:** Automated **Blinding:** Single; independent observer of the PSG data **Data Quality Categorized:** NR **Criteria for Excluding Data:** NR **Qualifications:** NR

Evidence Table 12. Zamarron et al., 2003: Study Design, cont'd

Comparison of Results	Special Statistics
Definitions: • Apnea: absence of airflow >10 s • Hypopnea: reduction of airflow for at least 10s accompanied by a ≥ 4% decrease in the saturation of hemoglobin • Suspicion of OSA: periodogram shows a peak within the period 30 to 70 s in either SaO2 or heart rate signals; normal= absence of both signals • PSG AHI: hourly samples of sleep; AHI ≥ 10 = OSA **Combined SaO2 and Heart Rate:** • Sensitivity: 94% (89-97) • Specificity: 82% (75-88) • PPV: 87% (81-91) • NPV: 92% (85-96) • Accuracy: 89% (84-92) • Positive Likelihood Ratio: 5.35 (3.69 – 7.78) • Negative Likelihood Ratio: 0.07 (0.03 - 0.13) • Of 10 OSA false-negatives, 40% (n=8) had severe COPD and presented no periodogram peak in 30 to 70s for both SaO2 and heart rate and thusly classified negative by portable; 2 pts had AHIs very close to limit of 10 to 15; 50% were > 65 years of age or older • Of 12 false-positives, 30 to 50% had an AHI of slightly < 10 **Other Indices:** • $\underline{SaO_2}$: 41.3% normal, 58.7% abnormal; concordance between 2 observers was 0.90 (0.86 to 0.94) • Heart Rate: 47.6% normal; 52.4% abnormal; concordance between 2 observers was 0.91 (0.87 to 0.94) • Combined SaO_2 and Heart Rate: 49.7% normal; 50.3% abnormal; PPV: 87%(81-92); NPV: 92%(85-96);	**Bland-Altman Plot:** NR **ROC:** NR

Evidence Table 12. Zamarron et al., 2003: Study Design, cont'd

Data Loss/ Variability/Cost Benefit	Quality Indicators	Author's Conclusions/ Limitations
Data Loss: NR **Night to Night Variability:** NR **Cost Benefit Analyses:** As a screening tool for the diagnosis of OSA, pulse oximetry is cost effective and shows substantial accuracy	**Evidence Level:** • Subjects selected consecutively or randomly (no design influence): N • Both tests scored blindly: Y • Reference standard done on all subjects: Y **Quality Score:** • Prospective study: Y • Portable done outside the lab: N • Random order of allocation of subjects to PSG or Portable first: N • Low data loss $\leq$ 10% of flow/SaO_2/EEG as applicable: N • High % completed ($\geq$ 90% of those entered): Y • PSG methodology fully described: Y • Portable monitor methodology fully described: Y • Portable scoring fully described: Y **Evidence Level Score:** II **Quality Level Score:** C **Grade:** Fair	**Conclusions:** • If the test result is negative it is unlikely that the patient will receive an OSA diagnosis • 30-70 s peaks in the periodograms in either signal had a sensitivity of 94% and 82% specificity, a PPV of 87% and NPV of 92% with respect to OSA diagnosis **Limitations:** Study not designed for upper airway resistance syndrome which presents no modifications of oxygen saturation but is associated with arousals related to respiratory effort (3 patients in study had UARS)