

Obstructive Sleep Apnea Devices for Out-Of-Center (OOC) Testing: Technology Evaluation

Nancy A. Collop, M.D.¹; Sharon L. Tracy, Ph.D.²; Vishesh Kapur, M.D.³; Reena Mehra, M.D., M.S.⁴; David Kuhlmann, M.D.⁵; Sam A. Fleishman, M.D.⁶; Joseph M. Ojile, M.D.⁷

¹Emory Sleep Center, Atlanta, GA; ²The American Academy of Sleep Medicine, Darien, IL; ³University of Washington Medicine Sleep Center, University of Washington, Seattle, WA; ⁴Case Western Reserve University, Cleveland, OH; ⁵Bothwell Regional Health Center, Sleep Disorders Center, Sedalia, MO; ⁶Cape Fear Valley Sleep Medicine Center, Fayetteville, NC; ⁷St. Louis University and Clayton Sleep Institute, Saint Louis, MO

Guidance is needed to help clinicians decide which out-of-center (OOC) testing devices are appropriate for diagnosing obstructive sleep apnea (OSA). A new classification system that details the type of signals measured by these devices is presented. This proposed system categorizes OOC devices based on measurements of Sleep, Cardiovascular, Oximetry, Position, Effort, and Respiratory (SCOPER) parameters. Criteria for evaluating the devices are also presented, which were generated from chosen pre-test and post-test probabilities. These criteria state that in patients with a high pre-test probability of having OSA, the OOC testing device has a positive likelihood ratio (LR+) of 5 or greater coinciding with an in-lab-polysomnography (PSG)-generated apnea hypopnea index (AHI) ≥ 5 , and an adequate sensitivity (at least 0.825). Since oximetry is a mandatory signal for scoring AHI using PSG, devices that do not incorporate oximetry were excluded. English peer-reviewed literature on FDA-approved devices utilizing more than 1 signal was reviewed according to the above criteria for 6 questions. These questions specifically addressed the adequacy of different respiratory and effort sensors and combinations thereof to diagnose OSA. In summary, the literature is currently inadequate to state with confidence that a thermistor alone without any effort sensor is adequate to diagnose OSA; if a thermal sensing device is used as the only measure of respiration, 2 effort belts are required as part of the montage and piezoelectric belts are acceptable in this context; nasal pres-

sure can be an adequate measurement of respiration with no effort measure with the caveat that this may be device specific; nasal pressure may be used in combination with either 2 piezoelectric or respiratory inductance plethysmographic (RIP) belts (but not 1 piezoelectric belt); and there is insufficient evidence to state that both nasal pressure and thermistor are required to adequately diagnose OSA. With respect to alternative devices for diagnosing OSA, the data indicate that peripheral arterial tonometry (PAT) devices are adequate for the proposed use; the device based on cardiac signals shows promise, but more study is required as it has not been tested in the home setting; for the device based on end-tidal CO₂ (ETCO₂), it appears to be adequate for a hospital population; and for devices utilizing acoustic signals, the data are insufficient to determine whether the use of acoustic signals with other signals as a substitute for airflow is adequate to diagnose OSA. Standardized research is needed on OOC devices that report LR+ at the appropriate AHI (≥ 5) and scored according to the recommended definitions, while using appropriate research reporting and methodology to minimize bias.

Keywords: Obstructive sleep apnea, out-of-center testing, SCOPER, testing devices, AHI

Citation: Collop NA; Tracy SL; Kapur V; Mehra R; Kuhlmann D; Fleishman SA; Ojile JM. Obstructive sleep apnea devices for out-of-center (OOC) testing: technology evaluation. *J Clin Sleep Med* 2011;7(5):531-548.

TABLE OF CONTENTS

Abstract	531
1.0 Introduction	532
2.0 Determining Criteria for Evaluating Devices	533
3.0 The SCOPER Categorization System	534
3.1 Development	534
3.2 Assessment of each SCOPER category	534
4.0 Classification of OOC Devices by SCOPER	534
5.0 Methods and Key Questions	536
6.0 Results	536
6.1 Key Question 1	536
6.2 Key Question 2	536
6.3 Key Question 3	537
6.4 Key Question 4	538
6.5 Key Question 5	538
6.6 Key Question 6	538

6.6.1 PAT Signal	539
6.6.2 Cardiac Signals plus oximetry	540
6.6.3 ETCO ₂ as an Alternate Measure of Airflow	540
6.6.4 Acoustic Signals as a Substitute for Airflow	540
7.0 Conclusions	541
8.0 Future Directions	541
References	542
Acknowledgments	543
Appendix I: Methods	544
Search Strategy	544
Inclusion and Exclusion Criteria	544
Grading	544
Data Extraction.....	545
Appendix II: Excluded Literature	547
Appendix III: Additional Outcomes Information on Devices from Excluded Literature	548

1.0 INTRODUCTION

The first widely used classification system for describing sleep testing devices was published by the American Academy of Sleep Medicine (then the American Sleep Disorders Association) in 1994, placing available devices into 4 categories based upon the number and type of “leads” used and the circumstances in which the device was used. This schema closely mirrored available Current Procedural Terminology (CPT) codes, worked for the majority of the then-available devices, and served to foster development of practice guidelines and reimbursement decisions. However, since that time, a plethora of innovative testing devices have been developed, many of which do not fit well within that classification scheme. In 2010, the Board of the American Academy of Sleep Medicine (AASM) commissioned a task force to determine a more specific and inclusive method of classifying and evaluating sleep testing devices other than polysomnography (PSG) used as aids in the diagnosis of obstructive sleep apnea (OSA) in the out-of-center (OOC) setting. The scope of this work was specifically limited to classification and evaluation of the performance characteristics of the technology itself and will not address their use in practice guidelines, accreditation standards, or management principles.

There are many issues involved in classifying and evaluating the performance characteristics of the wide array of devices purporting to diagnose OSA outside of the realm of attended polysomnography: (1) many different sensors might be used to measure the same physiologic parameter; (2) sensors may be combined in varied ways in an effort to enhance accuracy; (3) different physiologic parameters might be measured in one device vs. another; (4) signals may be modified by analog or digital processing to arrive at derived measures; and (5) studies evaluating and comparing devices may have employed varied “gold standards” or outcome measures. These and other factors make comparisons and generalizations between studies of even similar devices difficult.

The overarching purpose of our technology assessment is to provide a means of answering some pertinent clinical questions:

1. Is a thermal sensing device without an effort measure adequate to diagnose OSA?
2. Is a thermal sensing device with a measure of effort adequate to diagnose OSA?
3. Is nasal pressure without an effort measure adequate to diagnose OSA?
4. Is nasal pressure with an effort measure adequate to diagnose OSA?
5. With an effort measure, is nasal pressure in combination with a thermal sensing device significantly better than either a thermal sensing device or nasal pressure alone to warrant the requirement of both sensors?
6. What is the evidence for alternative devices to diagnose OSA?

We will first discuss an approach to addressing the complexities and challenges involved in assessing the performance and characteristics of OOC devices, and then use the methods devised to address the clinical questions above.

Variations in the Standards and Outcomes Used to Evaluate and Compare Devices

In addition to variability in sensors used in PSG, the definitions for apneas, hypopneas, apnea-hypopnea index (AHI), respiratory disturbance index (RDI), and OSA are also variable. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*¹ was written in an attempt to standardize the scoring definitions, but even the manual has 2 acceptable definitions of hypopnea. To emphasize the impact that different hypopnea scoring has on resultant AHI, Ruehland et al.² reported that the AHI of studies scored utilizing the “Chicago” criteria (hypopnea defined by a 50% reduction in airflow or < 50% reduction in airflow associated with a 3% oxygen desaturation and/or an arousal) was **3 times** the AHI of the same studies scored using the recommended hypopnea definition from the AASM scoring manual. To further complicate matters, the scoring used for an OOC device may or may not be the same as that used for the in-lab PSG because different sensors are used (including lack of a standard sleep measurement). Additionally, the definition of what level of AHI is considered adequate to confirm a diagnosis of OSA that qualifies for treatment is variable across studies. In this paper, we will use the definition of OSA-positive as an AHI ≥ 5 .

For this document, the following definitions will apply, except as used by the study’s author in which case their terminology was used:

The standard definition of AHI (AHI_s) as determined during attended laboratory PSG is:

$$\text{AHI}^s = \frac{[\text{apneas (10 sec without flow)} + \text{hypopneas (reduced flow with 4\% desat)}]}{\text{Total sleep time (h)}}$$

The nonstandard definition of AHI (AHI_{ns}) is defined as:

$$\text{AHI}^{ns} = \frac{[\text{all PSG determined respiratory events (apneas, hypopneas using other definitions, RERAs)]}}{\text{Total sleep time (h)}}$$

The respiratory event index (REI) is defined in the context of OOC testing devices as:

$$\text{REI} = \frac{[\text{apneas} + \text{hypopneas}]}{\text{Total sleep or recording time (h)}}$$

In addition to obtaining physiologic measures, many studies examining OOC devices report a variety of additional outcomes measures, such as compliance with positive airway pressure or change in subjective sleepiness. This makes summarizing the evidence across studies exceedingly challenging.

Keeping these challenges in mind, we have attempted to account for these challenges in 2 ways.

1. We have translated the varied outcome measures of studies evaluating portable testing devices into a dimensionless, useful parameter—the positive likelihood ratio (LR+) delivered by applying a given test and obtaining a “positive” result. This allows comparisons across a wide variety of devices that is less sensitive to variations in case definitions.

- We have developed a device categorization scheme that is adaptable, descriptive, and we believe workably specific (see SCOPER system below).

2.0 DETERMINING CRITERIA FOR EVALUATING DEVICES

Criteria need to be established to evaluate the appropriateness of OOC devices to diagnose OSA. The 2007 Clinical Guidelines³ prescribe that OOC devices should be used in patients with a “high pretest probability” for OSA. The following section outlines the logic used in establishing the definition of “high pretest probability.” The clinical determination of the pre-test probability is beyond the scope of this document, but will be addressed in a companion paper.

Essentially, the OOC device should be used to increase the pretest probability to a sufficiently high post-test probability that one is very certain that the patient has OSA. For the purposes of this paper, we will recommend that to be considered as having OSA, the post-test probability should be $\geq 95\%$. The relationship between the pre- and post-test probability may be described by the likelihood ratio, or LR. We are most concerned with using OOC devices to “rule in” OSA, and therefore we are interested in the positive likelihood ratio, LR+.⁴ The criteria for using OOC devices to “rule out” OSA may be different and are not addressed here.

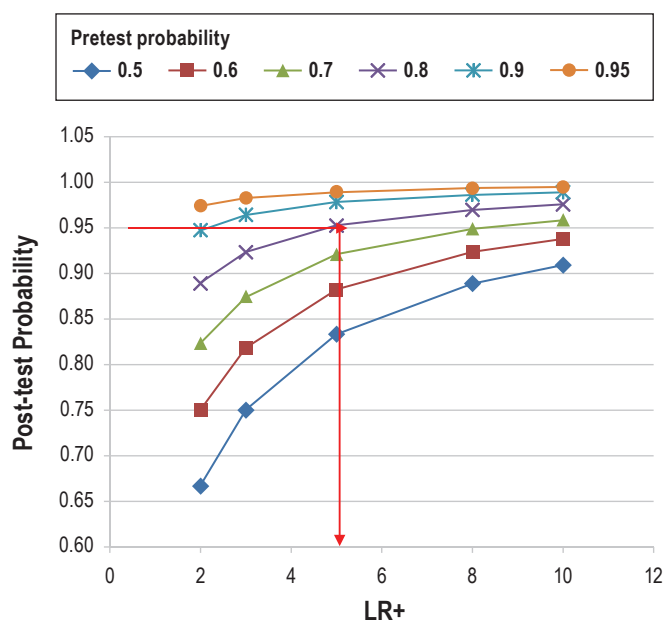
The combination of the following 3 equations describes the relationship between the LR+, pretest probability, and post-test probability and is shown graphically in **Figure 1**:

- $\text{odds}_{\text{pre}} = \text{probability}_{\text{pre}} / (1 - \text{probability}_{\text{pre}})$
- $\text{odds}_{\text{post}} = \text{odds}_{\text{pre}} \times \text{LR+}$
- $\text{probability}_{\text{post}} = \text{odds}_{\text{post}} / (1 + \text{odds}_{\text{post}})$

This foundation enables a judgment to be made as to a device’s adequacy to be used to help diagnose OSA given a patient’s specific pretest probability. **Figure 1** shows the relationship between post-test probability and LR+ at a variety of pretest probabilities between 0.5 and 0.95. The required device LR+ to achieve a post-test probability of 95% increases as the pretest probability decreases. For example, if the pretest probability is only 50%, the required LR+ would be greater than 20 (which is off the scale on this figure). For the purposes of this paper, a minimum pretest probability of 80% is suggested such that a reasonable requirement for LR+ can be employed for assessing devices. From **Figure 1**, at the given pretest (80%) and post-test (95%) probabilities, the device must have an LR+ of at least 5 to be clinically useful. This LR+ cutoff value will change if different pretest or post-test probabilities are employed. Similarly, if the device has an LR+ greater than 5, it could be acceptable at lower pretest probabilities. Also, if the device has an LR+ of 5 and the population has a pretest probability of only 50%, the post-test probability of disease drops to only 83%.

We also wanted to be sure that we are not increasing the burden of testing to the patients. That is, will the burden of some patients getting 2 tests (complex and simple) outweigh

Figure 1—The relationship between LR+, pretest probability, and post-test probability

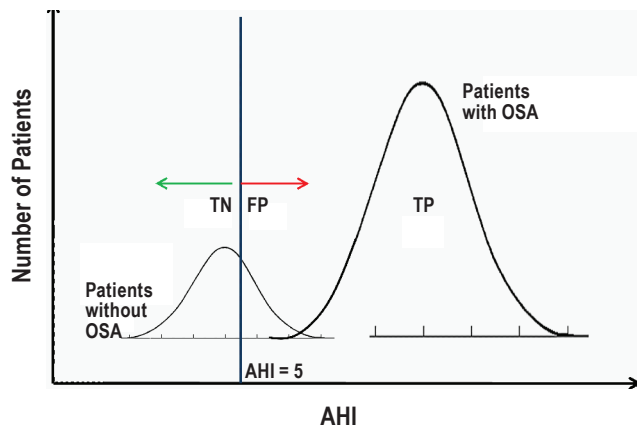


the benefit of some patients needing only 1 simpler test? This depends on how much burden the simple test causes relative to the gold standard as well as other factors (e.g., patient doesn’t follow up after initial negative test). We suggest that we want to have at least two-thirds of the population (66%) be diagnosed accurately as positive with the simple test; therefore, we set a minimum value for the sensitivity at 0.825. We have also included in the results (**Tables 3-11**) the negative likelihood ratios, but have concentrated on the positive LR in our schema.

Additionally, the definition of what level of AHI is considered adequate to confirm a diagnosis of OSA is variable across studies. According to the ICSD,⁵ an AHI ≥ 5 with symptoms is indicative of OSA. But should that definition remain the same for OOC testing considering the variety of event definitions, event detection technologies, as well as a difference in clinical management protocol? The authors judged that it should, for reasons including the desire to minimize the number of follow-up in-lab PSGs. **Figure 2** illustrates the concept of setting the AHI cutoff at 5 to maximize true positives, true negatives, and minimize false negatives in the context of a high pretest probability for OSA. **To summarize, devices are judged on whether or not they can produce an LR+ of at least 5 and a sensitivity of at least 0.825 at an in-lab AHI of at least 5.**

Only if the authors defined OSA at a different cutoff or if they used the Chicago criteria⁶ (see footnote following article) were different cutoffs considered in this evaluation of LR+. As described previously, this is the proposed criterion because AHI utilizing the “Chicago criteria” results have been found to be roughly 3 times the 2007 Scoring Manual Recommended rule.^{1,2} Therefore an AHI of 15 determined using the 1999 rules would be roughly equivalent to an AHI of 5 using the current recommended rule.

Figure 2—Illustration of the combination of the populations of patients with and without OSA with respect to the AHI cutoff, high pretest probability, true positive, true negative, and false positive results



3.0 THE SCOPER CATEGORIZATION SYSTEM

3.1 Development

In 1994, the AASM (formerly the ASDA) published Practice Parameters for the Use of Portable Recording in the Assessment of Obstructive Sleep Apnea⁷ and an associated review paper⁸ that categorized out-of-center (portable) devices into 4 levels: (1) standard attended PSG; (2) comprehensive portable PSG (unattended); (3) modified portable sleep apnea testing (unattended, minimum of 4 channels including ventilation [at least 2 channels of respiratory movement or a combination of respiratory movement and airflow], heart rate or electrocardiography (ECG), and oxygen saturation; and (4) continuous single- or dual-bioparameter recording (unattended).⁸ However, it has become increasingly apparent that with the continual technological changes that occur over time, this categorization is no longer useful. Many devices do not fit into these categories. Therefore, a new categorization scheme is needed.

A new scheme is suggested based upon the sensors used to measure each of the following:

Sleep, **C**ardiovascular, **O**ximetry, **P**osition, **E**ffort, and **R**espiratory, i.e. **SCOPER**.

3.2 Assessment of Each SCOPER Category

Each category of **SCOPER** was assessed individually as described below:

1. **Sleep**: The presence of a measurement of sleep was not quantitatively evaluated. The logic employed is that the measurement of sleep relates in large part to the final assessment of the sleep disordered breathing index, i.e., whether the denominator of the index is **per hour of sleep** or **per hour of recording time**. This will predominantly affect the cutoffs of positive or negative diagnoses, which can be addressed with some calibration of the device to in-laboratory PSG studies. At least 1 study⁹ stated that the addition of sleep surrogate measurement

(e.g., actigraphy) did not improve the device's performance for patients with a high pretest probability for obstructive sleep apnea.

2. **Cardiovascular**: The cardiovascular measurement evaluation was focused on devices that used either the cardiac signal or a vascular signal (e.g., peripheral arterial tonometry) to derive a respiratory event index. This is the only signal in which the usual PSG signal (one lead of ECG) is "demoted" to a lower level as it is used on a PSG typically for rate and rhythm analysis, rather than assessment of disordered breathing. We realize that these signals often measure different physiologic signals (cardiac vs. vascular tone) but felt that the novel PAT signal fit best in this category, although one could also argue to put it in a different category (respiratory or sleep).
3. **Oximetry**: Since the definition of AHI as measured by conventional parameters relies on desaturation to identify many events, for the determination of REI, a device must include an oximeter.
4. **Position**: Although we felt important to note in the categorization, the presence of a measurement of position was not quantitatively evaluated since it is not routinely used to diagnose OSA. P_1 is considered video or visual confirmation of body position, and P_2 is considered any other method to determine body position that is non-visual in nature. The effect of positional variations on OSA is a topic for research.
5. **Effort**: The addition of a measurement of respiratory effort was included in the key questions that follow. The best effort measure as noted in the scoring manual¹ is respiratory inductance plethysmography (RIP) with 2 belts. Beyond this gold standard, research on the use of 1 RIP belt, 2 or 1 piezo belt and other effort measures is scanty with regards to OOC devices.
6. **Respiratory**: The evaluation focused on the measurement of airflow, either by conventional or alternative methods.

Levels of each category of **SCOPER** are outlined in **Table 1**. These levels are numbered from 1 up to 5 (depending on category) and are based on the type of sensor or measurement that the device uses for that category. Note that when the device does not measure a certain category, that category is not included in its **SCOPER** identification and a "0" is listed in that category in **Table 2**. Also, if the type of signal has not been adequately defined in the study to allow a number to be assigned, it is denoted by an "x". In particular, for oximetry, the sampling rates have typically not been described in the literature, which has led to most devices being designated " O_{1x} ", indicating a finger or ear oximeter where the sampling time and/or rate have not been adequately described.

4.0 CLASSIFICATION OF OOC DEVICES BY SCOPER

A systematic search of the literature was performed, and when peer-reviewed literature in English was available for an FDA-approved device, data were extracted according to standardized methodology (see **Appendix I**). These data were used to categorize the devices according to the **SCOPER** scheme as shown in **Table 2**. Devices that were used in more than 1

Table 1—SCOPER Categorization System

Sleep	Cardiovascular	Oximetry	Position	Effort	Respiratory
S ₁ – Sleep by 3 EEG channels* with EOG and chin EMG	C ₁ – more than 1 ECG lead – can derive events	O ₁ – Oximetry (finger or ear) with recommended sampling	P ₁ – Video or visual position measurement	E ₁ – 2 RIP belts	R ₁ – Nasal pressure and thermal device
S ₂ – Sleep by less than 3 EEG* with or without EOG or chin EMG	C ₂ – Peripheral arterial tonometry	O _{1x} – Oximetry (finger or ear) without recommended sampling (per Scoring Manual) or not described	P ₂ – Non-visual position measurement	E ₂ – 1 RIP belt	R ₂ – Nasal pressure
S ₃ – Sleep surrogate: e.g. actigraphy	C ₃ – Standard ECG measure (1 lead)	O ₂ – Oximetry with alternative site (e.g. forehead)		E ₃ – Derived effort (e.g. forehead versus pressure, FVP)	R ₃ – Thermal device
S ₄ – Other sleep measure	C ₄ – Derived pulse (typically from oximetry)	O ₃ – Other oximetry		E ₄ – Other effort measure (including piezo belts)	R ₄ – End-Tidal CO ₂ (ETCO ₂)
	C ₅ – Other cardiac measure				R ₅ – Other respiratory measure

Proper oximetry sampling is defined as 3 s averaging and a minimum of 10 Hz sampling rate (25 Hz desirable).¹*3 EEG channels defined as frontal, central and occipital. EEG, electroencephalography; EOG, electrooculography; EMG, electromyography; ECG, electrocardiography; RIP, respiratory inductance plethysmography.

Table 2—Devices by SCOPER Categorization

Device Name	Sleep	Cardiac	Oximetry	Position	Effort	Respiratory
ApneaLink (Ng 2009)	0	4	1x	0	0	2
Apnoescreen I (Golpe 2002)	3	4	1x	2	0	3
Apnoescreen II (Garcia-Diaz 2007)	3	3	1x	2	4	3
ARES (Westbrook 2005)	3	4	2	2	0	5
ARES (Ayappa 2008, To 2009)	3	4	2	2	3	2
Compumedics PS-2 (Iber 2004)	2	3	1x	0	1	3
Embletta PDS (Ng 2010)	0	4	1x	2	1	2
Embletta (Dingli 2003)	0	0	1x	2	4	2
Morpheus Hx software with standard hospital signals (Amir 2010)	4	3	1x	0	4	4
Northeast Monitoring Holter-oximeter (Heneghan 2008)	0	3	1x	0	0	0
Novasom QSG/Bedbugg/Silent Night (Reichert 2003)	0	4	1x	0	x	5
Novasom QSG/Bedbugg/Silent Night (Claman 2001)	0	4	1x	0	4	5
Remmers/SnoreSat (Jobin 2007)	0	0	1x	2	0	5
Siesta (Campbell 2010)	2	3	1x	2	4	1
SNAP (Michaelson 2006)	0	4	1x	0	0	5
SNAP (Su 2004)	0	4	1x	0	x	5
Somt�/Morpheus (Takama 2010)	0	4	1x	0	x	3
Stardust II (Yin 2006, Santos-Silva 2009)	0	4	1x	2	4	2
WatchPAT (Bar 2003)	0	2	1x	2	0	0
WatchPAT (Ayas 2003, Pittman 2004, Pittman 2006, Zou 2006, Pang 2007, Choi 2010)	3	2	1x	2	0	0

Table 3—Device with thermistor as the only measure of respiration (R_3)

SCOPER/Device/Author (year)	Evidence Level	Setting	LR+	LR-	Sensitivity
			at REI/AHI _{ns} ≥ 10		
S ₃ C ₄ O _{1x} P ₂ R ₃ /Apnoescreen I/Golpe et al. (2002)	IIb	H/L	2.7	0.31	0.78

configuration have more than 1 SCOPER categorization. We limited the categorization to those configurations with appropriate literature and did not list all the configurations possible per manufacturers' specifications. We have not included (1) FDA-approved devices for which there is no literature; (2) non-FDA approved devices that do not have a current related FDA-approved device on the market; (3) single-channel devices; (4) therapeutic devices used in diagnostic mode; or (5) devices without oximeters.

5.0 METHODS AND KEY QUESTIONS

We developed a series of questions to evaluate the OOC devices. Because oxygen desaturation is currently required for scoring certain defined sleep related breathing events, acceptable devices must have oximetry as a mandatory signal. Therefore any device without an oximeter is currently not considered an acceptable device. Subsequently, all devices reported on herein have an oximeter. Because the effort and respiratory categories of SCOPER are significant parameters for diagnosing OSA, they will be the focus of this evaluation, which addresses the following key questions:

1. Is a thermal sensing device without an effort measure adequate to diagnose OSA?
2. Is a thermal sensing device with a measure of effort adequate to diagnose OSA?
3. Is nasal pressure without an effort measure adequate to diagnose OSA?
4. Is nasal pressure with an effort measure adequate to diagnose OSA?
5. With an effort measure, is nasal pressure in combination with a thermal sensing device significantly better than either a thermal sensing device or nasal pressure alone to warrant the requirement of both sensors?
6. What is the evidence for alternative devices to diagnose OSA?

Details on the methodology used to find, grade, and extract the literature to answer these questions are described in **Appendix I** (pages 544-6), as well as the grading results for each study. In particular, the setting of the tests is considered important. The best evidence for an OOC device is when tested concurrently with PSG and also tested OOC (designated as "home/lab" or H/L). The studies designated as H/H ("home/home") denotes the device was tested in the home against a comprehensive portable PSG device also performed in the home; and L/L ("lab/lab"), where the device was only tested simultaneously against PSG in the lab. The sensitivity of the devices was also reported in order to show the rate of false negatives. In some cases, the LR+ was calculated from the data in the paper based on our defi-

nitions and what was reported by the authors. Lastly, although results at AHI/REI ≥ 5 were desired, often they were not presented at this cutoff value; in those cases the data are presented at the cutoff reported. Also ideally, the LR+ and sensitivity cutoffs should be met in the home setting, but because the data are so sparse, for this paper, if the criteria were met in any setting the device configuration described was considered acceptable.

Appendix II (page 547) lists the studies that were excluded with the reasons for their exclusion. **Appendix III** (page 548) lists other additional outcomes information that was found in the excluded literature that did not specifically provide data for this assessment.

6.0 RESULTS

6.1 Key Question 1

Is a thermal sensing device without an effort measure adequate to diagnose OSA?

There was 1 paper¹⁰ that contained data that could be used to address this question. **Table 3** summarizes the data. The Apnoescreen I, which did not have any effort measurement, did not produce an LR+ greater than 5 at the only reported REI/AHI_{ns} (≥ 10).

SUMMARY

The literature is currently inadequate to state with confidence that a thermal sensing device alone is adequate to diagnose OSA.

6.2 Key Question 2

Is a thermal sensing device with a measure of effort adequate to diagnose OSA?

Three papers addressed this question. The data are summarized in **Table 4**. Two of the papers compared 2 OOC devices to each other: in Iber et al.,¹¹ the same device (with a thermocouple) was used both in the home and in the lab (S₂C₃O_{1x}E₁R₃/PS-2 System), and in Takama and Kurabayashi,¹² a simpler device (C₄O_{1x}E_xR₃/Somté Morpheus) was compared to a more complex one (S₂C₄O_{1x}E₁R₃/P-Series Plus) both with thermistors in the laboratory. Interestingly, the direct comparison of the same device in different environments (Iber et al.) provided an LR+ of only 3.1. The authors stated that their observed differences were equivalent to the variability of repeated studies in the same setting. Positional variation was a reason posed for the variability, but unfortunately, their device did not measure position so a detailed analysis was not possible. In Takama and Kurabayashi's study, an LR+ of 5.8

Table 4—Devices with thermistor as the only measure of respiration (R_3) with effort measurement

SCOPER/Device/ Author (year)	Evidence Level	Setting	LR+	LR-	Sensitivity
$S_3C_3O_{1x}P_2E_4R_3$ / Apnoescreen II/Garcia- Díaz et al. (2007)	Ia	L/L	7.9 to 24 at REI/AHI _{ns} ≥ 10*	0.06 at REI/AHI _{ns} ≥ 10*	0.95 at REI/AHI _{ns} ≥ 10
	Ia	H/L	10.5 to ∞ at REI/AHI _{ns} ≥ 10*	0.14 to 0.18 at REI/AHI _{ns} ≥ 10*	0.84-0.86 at REI/AHI _{ns} ≥ 10*
$S_2C_3O_{1x}E_1R_3$ /PS-2 System/Iber et al. (2004)	Ila	H/L	3.1** at REI ≥ 13.81	0.77** at REI ≥ 13.81	0.31** at REI ≥ 13.81
$C_4O_{1x}E_1R_3$ vs. $S_2C_4O_{1x}E_1R_3$ /Somté/ Morpheus vs. P-Series Plus/Takama and Kurabayashi (2010)	IVb	L/L	5.8 at REI/AHI _s ≥ 20	0.22 at REI/AHI _s ≥ 20	0.81 at REI/AHI _s ≥ 20

*This study compared 2 scorers (reason for the range in LR+). **Calculated from reported data.

Table 5—Devices using only nasal pressure as a measurement of respiration (R_2)

SCOPER/Device/ Author (year)	Evidence Level	Setting	LR+	LR-	Sensitivity
$C_4O_{1x}R_2$ /ApneaLink/ Ng et al. (2009)	Ia	L/L	∞ for all REI/AHI _{ns} (5-20)	0 for all REI/AHI _{ns} (5-20)	1.0 at REI/AHI _{ns} ≥ 5
$S_3C_4O_2P_2E_3R_2$ /ARES/ Ayappa et al. (2008)	Ia	L/L	6.0* at REI/AHI _s ≥ 5	0.02 at REI/AHI _s ≥ 5	0.98 at REI/AHI _s ≥ 5
	Ia	H/L	4.4* at REI/AHI _s ≥ 5	0.12 at REI/AHI _s ≥ 5	0.90 at REI/AHI _s ≥ 5
$S_3C_4O_2P_2E_3R_2$ /ARES/ To et al. (2009)	Ila	L/L	∞ at REI/AHI _{ns} ≥ 5	0.11 at REI/AHI _{ns} ≥ 5	0.89 at REI/AHI _{ns} ≥ 5

*Defined as OSA+

was found for REI ≥ 20 for the simpler device in the home as compared to the complex one in the laboratory. No data were provided on lower REIs. This was only a Level IVb study because of lack of blinded scoring.

One other study⁹ compared a simpler device to laboratory PSG. The Apnoescreen II which uses a thermistor, produced an LR+ greater than 5 at both REI/AHI_{ns} reported (10 and 15). The Apnoescreen II ($S_3C_3O_{1x}P_2E_4R_3$) used 2 piezoelectric effort belts, and it had an ECG lead. Two of the 3 studies met the sensitivity cutoff.

SUMMARY

This small data set indicates if a thermal sensing device is used as the only measure of respiration, 2 effort belts are required as part of the montage. It appears that piezoelectric belts are acceptable in this context.

6.3 Key Question 3

Is nasal pressure without an effort measure adequate to diagnose OSA?

Three papers¹³⁻¹⁵ addressed this question. The data are summarized in **Table 5**. The first device ($C_4O_{1x}R_2$) had an ad-

equated LR+ on 50 patients, but it has not been tested in the home.¹³ Further validation in the home setting is needed to confirm this result.

The other papers^{14,15} discuss the ARES device ($S_3C_4O_2P_2E_3R_2$). According to the manufacturer, the ARES relies on a change in nasal pressure and SpO₂, and confirms each event based on an arousal identified by changes in snoring, head movements, or sympathetic arousal (pulse rate increase). In the first study (Ayappa et al., Level Ia)¹⁴ of over 90 patients, the lab/lab portion of the study showed adequate LR+ (6.0 at REI/AHI_s ≥ 5), but the home/lab portion showed inadequate LR+ (4.4 at REI/AHI_s ≥ 5). It is acknowledged that meeting the LR+ criterion in the home setting is more challenging than in the lab setting. This evaluation is based on the AHIR4%_{ARES} vs AHI4%_{NPSG} data (other definitions and results were presented, but this definition most closely matches the Scoring Criteria). In a lab/lab comparison in the second study, To et al.¹⁵ (Level Ila) reported adequate results (∞ at AHI ≥ 5) in a large study on 141 patients.

SUMMARY

Although the 2 devices that fit into this Key Question have different configurations, they both show adequate LR+ and sensitivity values that indicate nasal pressure (with oximetry)

Table 6—Devices using nasal pressure (R₂) plus effort

SCOPER/Device/Author (year)	Evidence Level	Setting	LR+	LR-	Sensitivity
O _{1x} P ₂ E ₄ R ₂ /Embletta/Dingli et al. (2003) [2 Piezo belts]	Ia	H/L*	∞*	0.395	0.60*
C ₄ O _{1x} P ₂ E ₄ R ₂ /Embletta PDS/ Ng et al. (2010) [2 RIP belts]	Ib	L/L	6.5 at REI/AHI _{ns} ≥ 5	0.089 at REI/AHI _{ns} ≥ 5	0.92 at REI/AHI _{ns} ≥ 5
C ₄ O _{1x} P ₂ E ₄ R ₂ /Stardust III/ Santos-Silva et al. (2009) [1 piezo belt]	Ia	L/L	2.6 at REI/AHI _s ≥ 5	0.03 at REI/AHI _s ≥ 5	0.98 at REI/AHI _s ≥ 5
		H/L	2.5 at REI/AHI _s ≥ 5	0.08 at REI/AHI _s ≥ 5	0.95 at REI/AHI _s ≥ 5
C ₄ O _{1x} P ₂ E ₄ R ₂ /Stardust III/ Yin et al. (2006) [1 piezo belt]	IVa	H/L	1.0 at REI/AHI _s ≥ 5	N/A	1.00 at REI/AHI _s ≥ 5

*Calculated from reported data; at AHI_{ns} ≥ 15 and Embletta REI ≥ 20 defined as OSA+; L/L portion of study used to construct diagnostic cutoffs.

Table 7—Device using both thermistor and nasal pressure (R₁)

SCOPER/Device/Author (year)	Evidence Level	Setting	LR+	LR-	Sensitivity
S ₂ C ₃ O _{1x} P ₂ E ₄ R ₁ /Siesta/Campbell and Neill (2010)	Ia	H/L	1.8 at REI/AHI _{ns} ≥ 5	0.24 at REI/AHI _{ns} ≥ 5	0.88 at REI/AHI _{ns} ≥ 5

without an effort measure is adequate to diagnose OSA. However, the uniqueness of the ARES scoring may not make this configuration (nasal pressure alone without effort measure) broadly applicable, and further studies with this configuration are needed.

6.4 Key Question 4

Is nasal pressure with an effort measure adequate to diagnose OSA?

There were 4 papers that contained the required data to address this question. **Table 6** summarizes the data. One of them had adequate LR+ results (Ng et al.),¹⁶ and one had adequate LR+ results at an REI that had been previously calibrated to the in-lab results (Dingli et al.).¹⁷ Two studies reported inadequate LR+ (Yin et al.,¹⁸ Santos-Silva et al.¹⁹).

Dingli et al., who used 2 piezoelectric belts (O_{1x}P₂E₄R₂/Embletta) with a nasal pressure sensor, achieved a calculated LR+ of ∞ at AHI ≥ 15 on PSG, which was defined as OSA-positive. The lab/lab portion of the study was used to calibrate the device and cutoffs to in-lab results. For the LR+ calculation, the authors’ category of “possible OSA” [10 < AHI < 20] was added to the negative fraction for the results reported in **Table 6**. When the “possible OSA” category was recategorized as positive, the LR+ decreased to 4.

Two studies present LR+ results that were inadequate, and they were in the category C₄O_{1x}P₂E₄R₂. One piezoelectric belt was used to measure respiratory effort. Since the scoring criteria were equivalent or almost equivalent for the OOC and in-lab tests, it is unlikely that the inadequate LR+ results are due to scoring issues. The only time that the LR+ was adequate was at extremely high REIs (≥ 30 for Santos-Silva et al.¹⁹ and ≥ 50 for Yin¹⁸).

The sensitivity criterion was met in 3 of the 4 studies.

SUMMARY

The data indicate that nasal pressure can be an adequate measurement of respiration when either 2 piezoelectric or RIP belts are used to measure effort (but not 1 piezoelectric belt).

6.5 Key Question 5

With an effort measure, is nasal pressure in combination with a thermal sensing device significantly better than either a thermal sensing device or nasal pressure alone to warrant the requirement of both sensors?

There was only 1 Level Ia study²⁰ where both a nasal pressure transducer and a thermistor were used to report an LR+ outcome, and this device used 2 piezoelectric bands. The data, presented in **Table 7**, showed inadequate LR+ at REI/AHI ≥ 5. This is a counterintuitive result, and more data are needed. It is possible that more complex devices may not give better results.

SUMMARY

There is insufficient evidence to state that both nasal pressure and thermal sensing device are required to adequately diagnose OSA.

6.6 Key Question 6

What is the evidence for alternative devices to diagnose OSA?

The final question involves evaluating alternative devices that derive or calculate REI from signals other than those that directly measure respiration with either a thermistor and/or nasal pressure. Four alternate methods of determining REI were identified, based on the following signals:

1. Peripheral Arterial Tonometry (PAT)

Table 8—Devices using PAT signal (Watch PAT)

SCOPER/Device/Author (year)	Evidence Level	Setting	LR+	LR-	Sensitivity
C ₂ O _{1x} P ₂ /Bar et al. (2003)	Ia	L/L	7 at REI/AHI _{ns} ≥ 10*	0.33 at REI/AHI _{ns} ≥ 10*	0.7 at REI/AHI _{ns} ≥ 10*
S ₃ C ₂ O _{1x} P ₂ /Zou et al. (2006)	Ia	H/H	9 at REI/AHI _{ns} ≥ 10*	0.11 at REI/AHI _{ns} ≥ 10*	0.9 at REI/AHI _{ns} ≥ 10*
S ₃ C ₂ O _{1x} P ₂ /Pang et al. (2007)	Ia	L/L	4.7 at REI/AHI _{ns} ≥ 5	0.075 at REI/AHI _{ns} ≥ 5	0.94 at REI/AHI _{ns} ≥ 5
S ₃ C ₂ O _{1x} P ₂ /Pittman et al. (2004)	IIa	L/L	13.0 at REI/AHI _s ≥ 5	0 at REI/AHI _s ≥ 5	0.92 at REI/AHI _s ≥ 5
	IIa	H/L	∞ at REI/AHI _s ≥ 5	0 at REI/AHI _s ≥ 5	1.00 at REI/AHI _s ≥ 5
S ₃ C ₂ O _{1x} P ₂ /Pittman et al. (2006)	IIa	L/L	1.6 at REI/AHI _{ns} > 5†	0.29 at REI/AHI _{ns} > 5†	0.86 at REI/AHI _{ns} > 5†
S ₃ C ₂ O _{1x} P ₂ /Ayas et al. (2003)	IIa	L/L	2.9 at REI/AHI _{ns} ≥ 10‡	0.24 at REI/AHI _{ns} ≥ 10‡	0.83 at REI/AHI _{ns} ≥ 10‡
S ₃ C ₂ O _{1x} P ₂ /Choi et al. (2010)	IIb	L/Hospital	5.9 at REI/AHI _s ≥ 5	0 at REI/AHI _s ≥ 5	1 at REI/AHI _s ≥ 5

*Calculated from original figure in paper at AHI_{PSG} ≥ 10 and REI_{WatchPAT} scored according to Chicago criteria. †Scored according to Chicago criteria; if “converted” to standard criteria (see Section 1.0), at REI > 15 the LR+ is 8, which is adequate. ‡Scored according to Chicago criteria; if “converted” to standard criteria (see Section 1.0), at AHI ≥ 15 the LR+ is 3.5, which is also inadequate.

2. Cardiac signals plus oximetry
3. End-tidal carbon dioxide (ETCO₂) as an alternative measure of airflow
4. Acoustic signals as a substitute for airflow

The following is a summary of the data for the nonstandard methods of determining REI.

6.6.1 PAT Signal

There were 7 studies meeting inclusion criteria that compared WatchPAT (either S₃C₂O_{1x}P₂ or C₂O_{1x}P₂) to in-lab PSG. This device is based on the PAT signal, oximetry +/- actigraphy. There were 3 Level Ia studies, 3 Level IIa studies, and 1 Level IIb study on these devices. The data are summarized in **Table 8**.

The 3 Level Ia studies included Bar et al. (lab/lab portion),²¹ Zou et al.,²² and Pang et al.²³ In Bar et al., 102 patients included both healthy volunteers and those with suspected OSA. LR+ was calculated as 7 at REI/AHI_{ns} ≥ 10 (from Figure 7 in the original paper); REI was scored according to Chicago criteria, and the LR+ was adequate. Zou et al. performed a home/home study of 106 patients; the LR+ was calculated as 9 from Figure 4 in the original paper and was an adequate value. Pang et al. performed a lab/lab study on 37 patients with suspected OSA. The LR+ at REI/AHI_{ns} > 5 was 4.7, which is marginal.

The 3 Level IIa studies included Pittman et al.,²⁴ Pittman et al.,²⁵ and Ayas et al.²⁶ Pittman et al.²⁴ included both lab/lab and home/lab comparisons in this study on 30 patients. REIs were presented for the in-lab PSG using both the Chicago and standard AHI criteria and compared to REIs generated with a proprietary algorithm for the Watch PAT. The LR+ at REI ≥ 5 using standard AHI scoring criteria was 13.0 in the lab and ∞ at home, which is adequate.

Pittman et al.²⁵ assessed residual SDB during CPAP therapy with the Watch PAT in a lab/lab study. Using the Chicago criteria, the LR+ at REI > 5 was 1.6, which is inadequate. Ayas et al. performed a lab/lab study of 30 persons with and without

suspected OSA. For the AHI_{ns} threshold of 10 using Chicago criteria, the optimal LR+ (not defined at what REI cutoff) was 2.9, which is inadequate. For a threshold of AHI ≥ 15, it was also inadequate at LR+ = 3.5.

The remaining study was Level IIb (Choi et al.),²⁷ who reported the LR+ as 5.9 at REI/AHI_s ≥ 5 on 25 subjects studied in-lab and in the hospital; all lab PSGs were scored according to the 2007 scoring manual.

The sensitivity results were adequate for 6 of the 7 papers.

SUMMARY

Although the scoring criteria were variable and the results at PSG-AHI cutoff of 5 were not always reported, overall the data indicate that this device is adequate for the proposed use. Two of the 3 Level Ia studies reported adequate LR+ and one was marginal. One Level IIa study reported adequate LR+, 1 was marginal (depending on scoring and AHI cutoffs), and the other inadequate. The Level IIb study reported adequate LR+.

6.6.2 Cardiac Signals Plus Oximetry

One device (a Northeast Monitoring Holter-oximeter) based the REI on 1 ECG channel plus oximetry (C₃O_{1x}). The REI is calculated by a pattern recognition algorithm based on a combination of cyclic variations in heart rate associated with apnea, ECG-derived respiration, and SpO₂. In this Level Ia study by Heneghan et al.,²⁸ a lab/lab comparison was made using this device vs. PSG. The data are summarized in **Table 9**. The LR+ was adequate (∞ at REI/AHI_{ns} ≥ 5, 8.6 at REI/AHI_{ns} ≥ 10, 20.8 at REI/AHI_{ns} ≥ 15) from the 63 patients studied. Sensitivity was adequate for the REI/AHI_{ns} ≥ 5 criteria.

SUMMARY

Based on only one study of 63 patients, this device shows promise, but more study is required as it has not been tested in the home setting.

Table 9—Device using cardiac and oximetry signals

SCOPER/Device/Author (year)	Evidence Level	Setting	LR+	LR-	Sensitivity
C ₃ O _{1x} /Holter-oximeter/Heneghan et al. (2008)	Ia	L/L	∞ at REI/AHI _{ns} ≥ 5	0.06 at REI/AHI _{ns} ≥ 5	0.94 at REI/AHI _{ns} ≥ 5

Table 10—Device using ETCO₂

SCOPER/Device/Author (year)	Evidence Level	Setting	LR+	LR-	Sensitivity
S ₄ C ₃ O _{1x} E ₄ R ₄ /Morpheus Hx software/Amir (2010)	IIb	L/L	16.5 at REI/AHI _s ≥ 5	0.03 at REI/AHI _s ≥ 5	0.97 at REI/AHI _s ≥ 5

Table 11—Devices using acoustic signals

SCOPER/Device/Author (year)	Evidence Level	Setting	LR+	LR-	Sensitivity
C ₄ O _{1x} E _x R ₅ /Novasom QSG/Reichert et al. (2003) [1 effort sensor]	Ia	L/L	10.6 at REI/AHI _{ns} ≥ 15	0.05 at REI/AHI _{ns} ≥ 15	0.95 at REI/AHI _{ns} ≥ 15
		H/L	5.4 at REI/AHI _{ns} ≥ 15	0.11 at REI/AHI _{ns} ≥ 15	0.91 at REI/AHI _{ns} ≥ 15
C ₄ O _{1x} E ₄ R ₅ /Novasom QSG/Claman et al. (2001) [1 effort sensor]	Iva	L/L	14 at REI/AHI _{ns} ≥ 15	0.15 at REI/AHI _{ns} ≥ 15	0.86 at REI/AHI _{ns} ≥ 15
C ₄ O _{1x} E _x R ₅ /SNAP/Su et al. (2004) [1 effort sensor]	Ia	L/L	1.6 at REI/AHI _{ns} ≥ 5	0.05 at REI/AHI _{ns} ≥ 5	0.98 at REI/AHI _{ns} ≥ 5
C ₄ O _{1x} R ₅ /SNAP/Michaelson et al. (2006)	Ila	L/L	7.1 [†] at REI/AHI _{ns} ≥ 5	0.07 [†] at REI/AHI _{ns} ≥ 5	0.94 at REI/AHI _{ns} ≥ 5
O _{1x} P ₂ R ₅ /Remmers/Snoresat/Jobin et al. (2007)	Ia	H/H	3.68-4.42* at REI/AHI _{ns} ≥ 5	0.08-0.31* at REI/AHI _{ns} ≥ 5	0.74-0.94* at REI/AHI _{ns} ≥ 5
S ₃ C ₄ O ₂ P ₂ R ₅ /ARES/Westbrook et al. (2005)	Ia	L/L	6.8 at REI/AHI _{ns} > 10	0.03 at REI/AHI _{ns} > 10	0.974 at REI/AHI _{ns} > 10
		H/L	2.3 at REI/AHI _{ns} > 5	0.07 at REI/AHI _{ns} > 5	0.957 at REI/AHI _{ns} > 5

[†]These results are from the “PSG2” lab that scored using the recommended scoring rules. *Depends on the Suzanne recorder scoring rules.

6.6.3 ETCO₂ as an Alternate Measure of Airflow

Amir et al.²⁹ studied Morpheus Hx software with the use of standard hospital signals including 1 ECG channel, pulse oximetry, 2 piezoelectric chest and abdominal effort straps (S₄C₃O_{1x}E₄R₄), and ETCO₂. In this Level IIb study, the LR+ was adequate (16.5 at REI/AHI_s ≥ 5 and 13.7 at REI/AHI_s ≥ 15). Sensitivity was adequate as well. **Table 10** summarizes the data.

SUMMARY

Available data addresses only a hospital population and not a home-based population. It appears to be adequate in this situation.

6.6.4 Acoustic Signals as a Substitute for Airflow

There were a variety of devices that used acoustic signals with or without other signals as a substitute for airflow to characterize sleep disordered breathing. The data are summarized in **Table 11**.

There were 2 studies, 1 Level Ia (Reichert et al.)³⁰ and 1 Level IVa (Claman et al., Level IV because blinding not stated),³¹

on the device Novasom QSG (C₄O_{1x}E_xR₅ and C₄O_{1x}E₄R₅, respectively; also known as Bedbugg or Silent Night). This device reports REI using an algorithm based on sound measurements obtained from 2 microphones at the upper lip that record snoring intensity and ambient noise. The device also has a finger oximeter and an effort measure (1 identified as a pressure transducer; the other was not identified). For REI/AHI_{ns} ≥ 10, Claman et al. reported an LR+ of 14 (adequate) in a lab/lab study. Reichert et al. reported an LR+ of 10.6 in a lab/lab comparison and 5.4 in a home/lab comparison at REI/AHI_{ns} ≥ 15. Data were not reported for lower AHI_s.

There were 2 studies on 2 versions of the SNAP (C₄O_{1x}E_xR₅ and C₄O_{1x}R₅). These devices are based on measures of acoustic oronasal airflow and analyzed using a proprietary algorithm to define REI. Su et al. (Level Ia study)³² used a version of the device that included 1 chest belt (type undefined), and reported an inadequate LR+ of 1.6 at REI/AHI_{ns} ≥ 5. Michaelson et al. (Level IIa study in a lab/lab comparison),³³ who did not report the use of any effort channel, reported an adequate LR+ (7.1) at REI/AHI_{ns} ≥ 5.

Another device used sound plus oxygen desaturation as an alternative method for determining REI. In a Level Ia study on Remmers/Snoresat ($O_{1x}P_2R_5$), Jobin et al.³⁴ compared the device-determined REI results to 3 different REIs obtained from the Suzanne recorder, $C_4O_{1x}P_2E_4R_2$ in a home/home study. Unfortunately, none of the compared definitions of REI were similar to that recommended in the scoring manual. Nonetheless, the reported LR+ at $REI/AHI_{ns} \geq 5$ ranged from 3.68 to 4.42, which are inadequate.

Finally, in a Level Ia study by Westbrook et al.³⁵ on an earlier version of the ARES device ($S_3C_4O_2P_2R_5$) that utilized quantitative acoustic scoring measures plus behavioral/autonomic arousal detection, the results were adequate for the lab/lab comparison (although only results at $AHI \geq 10$ were reported), but inadequate in the home setting, although it is acknowledged that achieving adequate results in the home setting is more challenging. The sensitivity results were adequate in both settings.

SUMMARY

For this alternative measure, the data are insufficient to determine whether the use of acoustic signals with other signals as a substitute for airflow is adequate to determine REI. This is because of the varied literature base where (1) only 1 study reported adequate LR+ at $REI/AHI_{ns} \geq 5$, but it was only a lab/lab study (Michaelson et al.)³³; (2) the reported results were at AHI cutoffs that were too high (Reichert et al., Claman et al.)^{30,31} and only performed lab/lab (Claman et al., Westbrook et al.)^{31,35}; and (3) the reported LR+ were too low at $REI/AHI_{ns} \geq 5$ (Su et al., Jobin et al., Westbrook et al. H/L).^{32,34,35}

Sensitivity was adequate for all studies in this category.

7.0 CONCLUSIONS

Due to the variety of OOC devices now available, the previously accepted method of categorization of these devices is unsuitable; therefore, a new classification system, SCOPER, was proposed in this paper. SCOPER will allow the easy classification of OOC devices based on the types of sensors that they use to aid in the diagnosis of OSA, including sleep, cardiac, oximetry, position, effort, and respiratory measures.

The results of this technology evaluation are that the literature is currently inadequate to state with confidence that a thermistor alone without any effort sensor is adequate to diagnose OSA. If a thermistor is used as the only measure of respiration, 2 effort belts are required as part of the montage. It appears that piezoelectric belts are acceptable in this context. The data indicate that nasal pressure can be an adequate measurement of respiration without an effort measure; however, at this time, this may be device specific and further research is required before recommending broad usage. Nasal pressure may also be used when either 2 piezoelectric or RIP belts are used to measure effort (but not 1 piezoelectric belt). There is insufficient evidence to state that both nasal pressure and thermistor are required to adequately diagnose OSA. With respect to alternative devices for diagnosing OSA, the data indicate that peripheral arterial tonometry (PAT) devices are adequate for the proposed use; the device based on cardiac signals shows promise, but more study is required as it has not been tested in the home setting; for the device based on end-tidal CO_2 (ETCO₂), it appears to

be adequate for a hospital population; and for devices utilizing acoustic signals, the data are insufficient to determine whether the use of acoustic signals with other signals as a substitute for airflow is adequate to diagnose OSA.

8.0 FUTURE DIRECTIONS

This paper is meant to be the first step in a comprehensive process to evaluate and subsequently make recommendations on how to use OOC testing devices in an outpatient population. It is anticipated that the next paper will address the important issues of determining pretest probability, interpreting study results, developing testing algorithms and treatment decisions.

In 2003, Flemons et al.³⁶ published an evidence review on the home diagnosis of sleep apnea. In that paper, the authors nicely outlined the types of parameters that should be followed to properly assess OOC testing devices. Unfortunately, in the 8 years since that publication, many studies lack the important information required to make useful comparisons. A more recent paper provides very specific details about performing research with these devices and should be referred to for more detail on this subject.³⁷ For the evaluation of OOC testing devices, future studies would greatly benefit by the use of consistent outcomes measures to facilitate direct comparisons and meta-analyses of studies. As described herein, at a minimum, LR+ at an REI/AHI_s cutoff of 5 along with sensitivity are the desired outcome measures.

Comparison to the gold standard in-lab PSG with respect to sensor selection, montage selection, and scoring methods would unify the data with respect to the device that is being compared. The recommended montage is listed in the scoring manual, and the companion paper defines the recommended method of scoring. Currently, the definition of AHI is based on the signal from an oximeter, hence the requirement for the presence of that sensor. Emerging technologies allowing respiratory events to be defined in novel ways will be welcomed innovations.

Systematically testing various sensors and combinations of sensors with OOC devices would help to answer the question of the minimum number and type of sensors that are required to acceptably diagnose patients with OSA in an unattended home setting. This was particularly noted in examining the effort signals in which there was often minimal or no data on the type of signal employed or whether 1 or 2 belts were utilized. In addition, the effect of measuring actual sleep time with various sensors including EEG, actigraphy, and other technologies versus recording time is an area to revisit in the future. This will allow for further refinement in the definition of REI as well.

In addition, following modern experimental designs to minimize bias (such as blinded scoring; prospective, randomized, controlled designs; assessing validated standard measures such as PSG on all patients; ensuring low data loss and high percentage of patients who complete studies; and fully describing the PSG and device sensors, montages, and scoring criteria) would enhance the level of confidence in the results. Funding from sources not invested in the results or an explanation of the role that the funding source played in directing the study and/or interpreting the results, would help address any conflicts of interest concerns. In addition, the use of experimental designs that simulate clinical use would be ideal, including the degree and detail of instruction given to patients. Obviously the sample

size is important, and a description of the sample with information about eligibility, dropouts, missing data, and refusal to participate should be included.

Other important issues relate to the study population; most studies have concentrated on white males without comorbidities, although some studies are beginning to branch out to more diverse populations, which include more females, other ethnic groups, and patients with comorbid diseases such as heart failure. In fact, 4 devices reported in 5 studies were tested only on patients with heart failure.³⁸⁻⁴² A related issue is the use of OOC devices to differentiate obstructive from central sleep apnea. This becomes more important when broader population groups are studied. Future iterations of the paper will address this in more detail as more information is available.

Another problematic issue in assessment of these studies includes the scoring of the OOC device. Many devices have proprietary algorithms which cannot necessarily be checked by the interpreting physician for accuracy. Again, clear discussion of the algorithm is mandatory, and the ability to review raw data is also required.

FOOTNOTE

These events must fulfill criterion 1 or 2, plus criterion 3 of the following:

1. A clear decrease (> 50%) from baseline in the amplitude of a valid measure of breathing during sleep. Baseline is defined as the mean amplitude of stable breathing and oxygenation in the 2 minutes preceding onset of the event (in individuals who have a stable breathing pattern during sleep) or the mean amplitude of the 3 largest breaths in the 2 minutes preceding onset of the event (in individuals without a stable breathing pattern).
2. A clear amplitude reduction of a validated measure of breathing during sleep that does not reach the above criterion but is associated with either an oxygen desaturation of > 3% or an arousal.
3. The event lasts 10 seconds or longer.

REFERENCES

1. Iber C, Ancoli-Israel S, Chesson A, Quan SF, for the American Academy of Sleep Medicine. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications, 1st ed.* Westchester, IL: American Academy of Sleep Medicine, 2007.
2. Ruehland WR, Rochford PD, O'Donoghue FJ, Pierce RJ, Singh P, Thornton AT. The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index. *Sleep* 2009;32:150-7.
3. Collop NA, Anderson WM, Boehlecke B, et al. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med* 2007;3:737-47.
4. Collop NA, Anderson WM, Boehlecke B, et al. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med* 2007;3:737-47.
5. American Academy of Sleep Medicine. *International classification of sleep disorders, 2nd ed.: Diagnostic and coding manual.* Westchester, IL: American Academy of Sleep Medicine, 2005.
6. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force. *Sleep* 1999;22:667-89.
7. Practice parameters for the use of portable recording in the assessment of obstructive sleep apnea. Standards of Practice Committee of the American Sleep Disorders Association. *Sleep* 1994;17:372-7.
8. 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.
9. Garcia-Diaz E, Quintana-Gallego E, Ruiz A, et al. Respiratory polygraphy with actigraphy in the diagnosis of sleep apnea-hypopnea syndrome. *Chest* 2007;131:725-32.
10. Golpe R, Jimenez A, Carpizo R. Home sleep studies in the assessment of sleep apnea/hypopnea syndrome. *Chest* 2002;122:1156-61.
11. Iber C, Redline S, Kaplan Gilpin AM, et al. Polysomnography performed in the unattended home versus the attended laboratory setting--Sleep Heart Health Study methodology. *Sleep* 2004;27:536-40.
12. Takama N, Kurabayashi M. Effectiveness of a portable device and the need for treatment of mild-to-moderate obstructive sleep-disordered breathing in patients with cardiovascular disease. *J Cardiol* 2010;56:73-8.
13. Ng SS, Chan TO, To KW, et al. Validation of a portable recording device (ApneaLink) for identifying patients with suspected obstructive sleep apnoea syndrome. *Intern Med J* 2009;39:757-62.
14. Ayappa I, Norman RG, Seelall V, Rapoport DM. Validation of a self-applied unattended monitor for sleep disordered breathing. *J Clin Sleep Med* 2008;4:26-37.
15. To KW, Chan WC, Chan TO, et al. Validation study of a portable monitoring device for identifying OSA in a symptomatic patient population. *Respirology* 2009;14:270-5.
16. Ng SSS, Chan T-O, To K-W, et al. Validation of Embletta portable diagnostic system for identifying patients with suspected obstructive sleep apnoea syndrome (OSAS). *Respirology* 2010;15:336-42.
17. 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.
18. Yin M, Miyazaki S, Ishikawa K. Evaluation of type 3 portable monitoring in unattended home setting for suspected sleep apnea: factors that may affect its accuracy. *Otolaryngol Head Neck Surg* 2006;134:204-9.
19. Santos-Silva R, Sartori DE, Truksinas V, et al. Validation of a portable monitoring system for the diagnosis of obstructive sleep apnea syndrome. *Sleep* 2009;32:629-36.
20. Campbell AJ, Neill AM. Home set-up polysomnography in the assessment of suspected obstructive sleep apnea. *J Sleep Res* 2011;20:207-13.
21. 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.
22. Zou D, Grote L, Peker Y, Lindblad U, Hedner J. Validation a portable monitoring device for sleep apnea diagnosis in a population based cohort using synchronized home polysomnography. *Sleep* 2006;29:367-74.
23. Pang KP, Gourin CG, Terris DJ. A comparison of polysomnography and the WatchPAT in the diagnosis of obstructive sleep apnea. *Otolaryngol Head Neck Surg* 2007;137:665-8.
24. Pittman SD, Ayas NT, MacDonald MM, Malhotra A, Fogel RB, White DP. Using a wrist-worn device based on peripheral arterial tonometry to diagnose obstructive sleep apnea: in-laboratory and ambulatory validation. *Sleep* 2004;27:923-33.
25. Pittman SD, Pillar G, Berry RB, Malhotra A, MacDonald MM, White DP. Follow-up assessment of CPAP efficacy in patients with obstructive sleep apnea using an ambulatory device based on peripheral arterial tonometry. *Sleep Breath* 2006;10:123-31.
26. 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.
27. Choi JH, Kim EJ, Kim YS, et al. Validation study of portable device for the diagnosis of obstructive sleep apnea according to the new AASM scoring criteria: Watch-PAT 100. *Acta Otolaryngol* 2010;130:838-43.
28. Heneghan C, Chua CP, Garvey JF, et al. A portable automated assessment tool for sleep apnea using a combined Holter-oximeter. *Sleep* 2008;31:1432-9.
29. Amir O, Barak-Shinar D, Amos Y, MacDonald M, Pittman S, White DP. An automated sleep-analysis system operated through a standard hospital monitor. *J Clin Sleep Med* 2010;6:59-63.
30. Reichert JA, Bloch DA, Cundiff E, Votteri BA. Comparison of the NovaSom QSGTM, a new sleep apnea home-diagnostic system, and polysomnography. *Sleep Med* 2003;4:213-8.
31. Claman D, Murr A, Trotter K. Clinical validation of the Bedbug in detection of obstructive sleep apnea. *Otolaryngol Head Neck Surg* 2001;125:227-30.
32. Su S, Baroddy FM, Kohrman M, Suskind D. A comparison of polysomnography and a portable home sleep study in the diagnosis of obstructive sleep apnea syndrome. *Otolaryngol Head Neck Surg* 2004;131:844-50.

33. Michaelson PG, Allan P, Chaney J, Mair EA. Validations of a portable home sleep study with twelve-lead polysomnography: comparisons and insights into a variable gold standard. *Ann Otol Rhinol Laryngol* 2006;115:802-9.
34. Jobin V, Mayer P, Bellemare F. Predictive value of automated oxygen saturation analysis for the diagnosis and treatment of obstructive sleep apnoea in a home-based setting. *Thorax* 2007;62:422-7.
35. Westbrook PR, Levendowski DJ, Cvetinovic M, et al. Description and validation of the apnea risk evaluation system: a novel method to diagnose sleep apnea-hypopnea in the home. *Chest* 2005;128:2166-75.
36. 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.
37. Kuna S, Badr MS, Kiimoff RJ, et al. An official ATS/AASM/ACCP/ERS workshop report: Research priorities in ambulatory management of adults with obstructive sleep apnea. *Proc Am Thorac Soc* 2011;8:1-16.
38. Abraham WT, Trupp RJ, Phillips B, et al. Validation and clinical utility of a simple in-home testing tool for sleep-disordered breathing and arrhythmias in heart failure: results of the Sleep Events, Arrhythmias, and Respiratory Analysis in Congestive Heart Failure (SEARCH) study. *Congest Heart Fail* 2006;12:241-7.
39. Abraham WT, Trupp RJ, Phillips B, et al. Effect of treatment with continuous positive airway pressure or oxygen on sleep-disordered breathing in patients with heart failure: results of the Sleep Events, Arrhythmias, and Respiratory Analysis in Chronic Heart Failure (SEARCH) study. *Congest Heart Fail* 2008;14:197-201.
40. Freimark D, Adler Y, Sheffy J, et al. Oscillations in peripheral arterial tone in congestive heart failure patients: a new marker for Cheyne-Stokes breathing. *Cardiology* 2002;98:21-4.
41. Quintana-Gallego E, Villa-Gil M, Carmona-Bernal C, et al. Home respiratory polygraphy for diagnosis of sleep-disordered breathing in heart failure. *Eur Respir J* 2004;24:443-8.
42. Smith LA, Chong DW, Vennelle M, Denvir MA, Newby DE, Douglas NJ. Diagnosis of sleep-disordered breathing in patients with chronic heart failure: evaluation of a portable limited sleep study system. *J Sleep Res* 2007;16:428-35.
43. Sackett DL, Strauss SE, Richardson WS, et al. *Evidence-based medicine: how to practice and teach EBM*. 2nd ed. Edinburgh, Scotland, UK: Churchill Livingstone, 2000.
44. Bagnato MC, Nery LE, Moura SM, Bittencourt LR, Tufik S. Comparison of Auto-Set and polysomnography for the detection of apnea-hypopnea events. *Braz J Med Biol Res* 2000;33:515-9.
45. BaHammam AS. Signal failure of type 2 comprehensive unattended sleep studies in patients with suspected respiratory sleep disordered breathing. *Sleep Breath* 2005;9:7-11.
46. Berry R, Hill G, Thompson L, McLaurin V. Portable monitoring and autotitration versus polysomnography for the diagnosis and treatment of sleep apnea. *Sleep* 2008;31:1423-31.
47. Bridevaux PO, Fitting JW, Fellrath JM, Aubert JD. Inter-observer agreement on apnoea hypopnoea index using portable monitoring of respiratory parameters. *Swiss Med Wkly* 2007;137:602-7.
48. Chua CP, Garvey J, Redmond S, Heneghan C, McNicholas WT. Towards automated sleep state estimation using a Holter-oximeter. *Conf Proc IEEE Eng Med Biol Soc* 2007:3998-4001.
49. Chung F, Liao P, Sun Y, et al. Perioperative practical experiences in using a level 2 portable polysomnography. *Sleep Breath* 2010; Mar 16. [Epub ahead of print].
50. Galer C, Yonkers A, Duff W, Heywood B. Clinical significance of SNAP somnography test acoustic recording. *Otolaryngol Head Neck Surg* 2007;136:241-5.
51. Gantner D, Ge J-Y, Li L-H, et al. Diagnostic accuracy of a questionnaire and simple home monitoring device in detecting obstructive sleep apnoea in a Chinese population at high cardiovascular risk. *Respirology* 2010;15:952-60.
52. Hedner J, Pillar G, Pittman SD, Zou D, Grote L, White DP. A novel adaptive wrist actigraphy algorithm for sleep-wake assessment in sleep apnea patients. *Sleep* 2004;27:1560-6.
53. Herscovici S, Pe'er A, Pappan S, Lavie P. Detecting REM sleep from the finger: an automatic REM sleep algorithm based on peripheral arterial tone (PAT) and actigraphy. *Physiol Meas* 2007;28:129-40.
54. Jurado Gámez B, Redel Montero J, Munoz Cabrera L, et al. [Cost-effectiveness and degree of satisfaction with home sleep monitoring in patients with symptoms of sleep apnea]. *Arch Bronconeumol* 2007;43:605-10.
55. Levendowski D, Steward D, Woodson BT, Olmstead R, Popovic D, Westbrook P. The impact of obstructive sleep apnea variability measured in-lab versus in-home on sample size calculations. *Int Arch Med* 2009;2:2.
56. Masdeu MJ, Ayappa I, Hwang D, Mooney AM, Rapoport DM. Impact of clinical assessment on use of data from unattended limited monitoring as opposed to full-in lab PSG in sleep disordered breathing. *J Clin Sleep Med* 2010;6:51-8.
57. Means MK, Edinger JD, Husain AM. CPAP compliance in sleep apnea patients with and without laboratory CPAP titration. *Sleep Breath* 2004;8:7-14.
58. Mueller A, Fietze I, Voelker R, et al. Screening for sleep-related breathing disorders by transthoracic impedance recording integrated into a Holter ECG system. *J Sleep Res* 2006;15:455-62.
59. Mulgrew A, Fox N, Ayas N, Ryan C. Diagnosis and initial management of obstructive sleep apnea without polysomnography. *Ann Intern Med* 2007;146:157-66.
60. Nakayama-Ashida Y, Takegami M, Chin K, et al. Sleep-disordered breathing in the usual lifestyle setting as detected with home monitoring in a population of working men in Japan. *Sleep* 2008;31:419-25.
61. Penzel T, Kesper K, Pinnow I, Becker HF, Vogelmeier C. Peripheral arterial tonometry, oximetry and actigraphy for ambulatory recording of sleep apnea. *Physiol Meas* 2004;25:1025-36.
62. Pépin J-L, Defaye P, Vincent E, Christophe-Boulard S, Tamisier R, Lévy P. Sleep apnea diagnosis using an ECG Holter device including a nasal pressure (NP) recording: validation of visual and automatic analysis of nasal pressure versus full polysomnography. *Sleep Med* 2009;10:651-6.
63. Pillar G. An automatic ambulatory device for detection of AASM defined arousals from sleep: the WP100. *Sleep Med* 2003;4:207-12.
64. Popovic D, King C, Guerrero M, Levendowski DJ, Henninger D, Westbrook PR. Validation of forehead venous pressure as a measure of respiratory effort for the diagnosis of sleep apnea. *J Clin Monit Comput* 2009;23:1-10.
65. Prasad B, Carley DW, Herdegen JJ. Continuous positive airway pressure device-based automated detection of obstructive sleep apnea compared to standard laboratory polysomnography. *Sleep Breath* 2010;14:101-7.
66. Quintana-Gallego E, Villa-Gil M, Carmona-Bernal C, et al. Home respiratory polygraphy for diagnosis of sleep-disordered breathing in heart failure. *Eur Respir J* 2004;24:443-8.
67. Skomro RP, Gjevre J, Reid J, et al. Outcomes of home-based diagnosis and treatment of obstructive sleep apnea. *Chest* 2010;138:257-63.
68. Smith LA, Chong DWS, Vennelle M, Denvir MA, Newby DE, Douglas NJ. Diagnosis of sleep-disordered breathing in patients with chronic heart failure: evaluation of a portable limited sleep study system. *J Sleep Res* 2007;16:428-35.
69. Torre-Bouscoulet L, Meza-Vargas M, Castorena-Maldonado A, Reyes-Zuniga M, Perez-Padilla R. Autoadjusting positive pressure trial in adults with sleep apnea assessed by a simplified diagnostic approach. *J Clin Sleep Med* 2008;4:341-7.
70. Whitelaw Wa, Brant RF, Flemons WW. Clinical usefulness of home oximetry compared with polysomnography for assessment of sleep apnea. *Am J Respir Crit Care Med* 2005;171:188-93.
71. Yin M, Miyazaki S, Itasaka Y, et al. A preliminary study on application of portable monitoring for diagnosis of obstructive sleep apnea. *Auris Nasus Larynx* 2005;32:151-6.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of the following members of the AASM staff: Christine Stepanski, M.S., for literature search contributions and help in coordinating the project; Kathleen McCann for help coordinating the review process; Sherene Thomas, Ph.D., for help with editing; Carolyn Winter-Rosenberg for help in coordinating the project; and Judy Coy, R.N., for help in formulating the project direction.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication August, 2011

Accepted for publication August, 2011

Address correspondence to: Sharon L. Tracy, Ph.D., American Academy of Sleep Medicine, 2510 North Frontage Road, Darien, IL 60561-1511; Tel: (630) 737-9700; Fax: (630) 737-9790; E-mail: stracy@aaasmnet.org

DISCLOSURE STATEMENT

This was not an industry supported study. Dr. Mehra has participated in paid speaking engagements. Dr. Tracy is an employee of the American Academy of Sleep Medicine. The other authors have indicated no financial conflicts of interest.

APPENDIX I: METHODS

Search Strategy

The search was performed in 2 parts. The first was a general search using PubMed of MEDLINE for relevant, original, peer-reviewed literature in English in the last decade (back to January 2000) and is shown in **Table A1**. The second consisted of device-specific searches in PubMed since its inception using the device names as search terms.

Device names were obtained in 4 ways: (1) by pearling the first literature search results for device names; (2) from FDA 510(k) device listings in the FDA database; (3) by literature obtained by company representatives at the SLEEP conference in 2010; and (4) from the September 2009 Sleep Review Magazine Home Sleep Testing Comparison Guide.

Inclusion and Exclusion Criteria

Only peer-reviewed English literature was included in the analyses, therefore data available from manufacturers on their websites, in product brochures, or published in non-peer-reviewed magazines were excluded. Only devices measuring 2 or more bioparameters were included.

Grading

The grading scheme was a modified form of that used in the 2003 “Home Diagnosis of Sleep Apnea: A Systematic Review of the Literature.”³⁶ The results of the evidence grading process are shown in **Table A2**.

Evidence level (I, II, III, and IV) (Table A3)

The presence or absence of three key indicators of quality dictated the assignment of evidence level based on an approach published by Sackett et al.⁴³ The definitions of these evidence levels are listed below as follows:

- I. blinded comparison, consecutive patients, reference standard 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;
- IV. reference standard was not applied blindly or independently.

The definitions of the three indicators used to assign level of evidence were as follows:

Blinded comparison: the portable monitor and polysomnogram were scored separately and without knowledge of the results of the other investigation; or the portable monitor study was scored automatically, and it was performed after the PSG was scored. If the investigators failed to mention whether or not the scorers were blinded, this criterion was deemed not to have been met.

Consecutive or random patients: the investigators did not participate in deciding what patients were included in the study. This criterion was met if patients were referred to a sleep *clinic* rather than a sleep *laboratory* (unless the investigators explicitly stated that they did not participate in selecting the patients referred to the laboratory). Either consecutively or randomly chosen patients were enrolled.

Reference standard was performed on all patients: all patients entered into the study must have undergone both a portable monitor test and a polysomnogram (either in-lab or comparison device (e.g. comprehensive portable PSG, depending on the aims of the study)). If the results of one test influenced the decision to perform the other, then this criterion was deemed not to have been met.

Quality rating (a, b, c, d) (Table A4)

Seven other aspects of a study’s methodology were scored, and a quality rating was assigned based on the number of indicators for which the study met the criteria. Al-

Table A1—Search Strategy and Search Terms

Column 1	Column 2	Column 3	Column 4	Column 5
Portable	Monitor*	Polysomnography[Mesh]	Sleep apnea syndromes [mesh]	Sleep Apnea Syndromes[Mesh]
Home	Care service*	PSG	Monitoring, ambulatory [majr]	Sleep Apnea
Limited	Sleep test*	Respiratory polygraphy		apnea
Ambulatory	Sleep study			
Unattended	Sleep studies			
Nonlaboratory	Screening tool			
In-home	Recording*			
	device			
	diagnosis			
	diagnostic			

((Column 1 AND Column 2) AND (Column 3 OR Column 4)) AND (Column 5)

Table A2—Results of Evidence Grading Process

Author, yr	SCOPER/Device	Evidence Level					Paper Quality					
		Blinded scoring	Consecutive or random pts	Full PSG on all pts?	SUMMARY	Prospective recruitment of pts	Random order of testing	Low data loss	High percentage completed	PSG methods/ definitions fully described	Device methods/ definitions fully described	Device scoring fully described
Golpe, 2002	S ₃ C ₀ P ₂ R ₂ /Aproscreen I	Y	N	Y	II	Y	N	N	Y	Y	Y	b
Garcia-Diaz, 2007	S ₃ C ₀ P ₂ E ₁ R ₂ /Aproscreen II	Y	Y	Y	I	Y	Y	Y	Y	N	Y	a
Iber, 2004	S ₂ C ₀ P ₂ E ₁ R ₂ /PS-2 System	Y	N	Y	II	Y	Y	N	Y	Y	Y	a
Takama, 2010	C ₀ P ₂ E ₁ R ₂ vs. S ₃ C ₀ P ₂ E ₁ R ₂ /Somté/Morpheus vs. P-Series Plus	N	Y	Y	IV	Y	N/A	N	Y	N	Y	b
Ng, 2009	C ₀ P ₂ E ₁ R ₂ /ApneaLink	Y	Y	Y	I	Y	N/A	Y	Y	Y	Y	a
Ayappa, 2008	S ₃ C ₀ P ₂ E ₁ R ₂ /ARES	Y	Y	Y	I	Y	N	L/L: Y H/L: N	Y	Y	Y	a
To, 2009		Y	N	Y	II	Y	N/A	N	Y	Y	Y	a
Dingli, 2003	O ₁ P ₂ E ₁ R ₂ /Embletta	Y	Y	Y	I	Y	Y	N	Y	Y	Y	a
Ng, 2010	C ₀ P ₂ E ₁ R ₂ /Embletta PDS	Y	Y	Y	I	Y	N/A	N	Y	N	Y	b
Santos-Silva, 2009	C ₀ P ₂ E ₁ R ₂ /Stardust II	Y	Y	Y	I	Y	Y	Y	Y	Y	Y	a
Yin, 2006		N	Y	N	IV	Y	N	Y	Y	Y	Y	a
Campbell, 2010	S ₃ C ₀ P ₂ E ₁ R ₂ /Siesta	Y	Y	Y	I	Y	Y	Y	Y	Y	Y	a
Bar, 2003	C ₀ P ₂ E ₁ R ₂ /WatchPat	Y	Y	Y	I	Y	N/A	Y	Y	Y	Y	a
Zou, 2006	S ₃ C ₀ P ₂ E ₁ R ₂ /WatchPat	Y	Y	Y	I	Y	N/A	Y	Y	Y	Y	a
Pang, 2007		Y	Y	Y	I	Y	N/A	N	Y	Y	Y	a
Pittman, 2004		Y	N	Y	II	Y	Y	Y	Y	Y	Y	a
Pittman, 2006		Y	N	Y	II	Y	N/A	Y	Y	Y	Y	a
Ayas, 2003		Y	N	Y	II	Y	N/A	Y	Y	Y	Y	a
Choi, 2010		Y*	N	Y	II	Y	N	Y	Y	Y	N	b
Heneghan, 2008	C ₀ P ₂ E ₁ R ₂ /Holter-oximeter	Y	Y	Y	I	Y	N/A	Y	Y	Y	Y	a
Amir, 2010	C ₀ P ₂ E ₁ R ₂ /Morpheus Hx software	Y	N	Y	II	Y	N/A	Y	Y	N	N	b
Reichert, 2003	C ₀ P ₂ E ₁ R ₂ /Novasom QSG	Y	Y	Y	I	Y	L: N/A H: N	Y	Y	Y	Y	a
Claman, 2001	C ₀ P ₂ E ₁ R ₂ /Novasom QSG	N (not stated)	Y	Y	IV	Y	N/A	Y	Y	Y	Y	a
Su, 2004	C ₀ P ₂ E ₁ R ₂ /SNAP	Y	Y	Y	I	Y	N/A	Y	Y	Y	N	a
Michaelson, 2006	C ₀ P ₂ E ₁ R ₂ /SNAP	Y	N	Y	II	Y	N/A	Y	Y	Y	N	a
Jobin, 2007	O ₁ P ₂ E ₁ R ₂ /Remmers/Snoresat	Y	Y	Y	I	Y	N/A	Y	Y	Y	Y	a
Westbrook, 2005	S ₃ C ₀ P ₂ E ₁ R ₂ /ARES	Y	Y	Y	I	Y	Y	Y	Y	Y	Y	a

*Blinded PSG scoring not explicitly stated, but PSGs scored before device studies, which were scored automatically, judged not to be a significant source of bias

Table A3

Evidence Level	Blinded scoring	Consecutive patients	PSG on all pts
I	Y	Y	Y
II	Y	N	Y
III	Y	Y	N
IV	PSG not applied blindly or independently		

though the random assignment of testing was an important indicator, it was not applicable to studies that had studied a portable monitor simultaneously with polysomnography. Thus, in some circumstances studies were rated on 6 indicators rather than 7. The quality indicator (a to d) was based on the number of indicators for which that study did not meet the criteria, as follows:

- a. zero or one quality indicators not met;
- b. two quality indicators not met;
- c. three quality indicators not met;
- d. four or more quality indicators not met.

The seven indicators and their definitions are listed below as follows:

1. Prospective recruitment of patients: the portable monitoring test and the polysomnogram were performed as patients were recruited into the study rather than reviewing a series of patients who had previously been studied.
2. Random order of testing: patients were assigned to undergo portable monitoring testing or polysomnography first at random rather than at the discretion of the investigators. If the portable monitoring study was performed simultaneously with the polysomnogram, this indicator was not rated.
3. Low data loss (< 10%): there were < 10% of patients whose results could not be compared because of the loss of polysomnography or portable monitoring data. This indicator allows for the repetition of studies to obtain acceptable results.
4. High percentage completed (> 90%): of the patients who were initially enrolled into the study (not counting a priori exclusions), > 90% completed the study protocol.
5. Polysomnography methodology/definitions fully described: the polysomnography methods must include the following:
 - a. characterization of the equipment used;

Table A4

Quality rating	
a	0 or one quality indicators not met
b	Two quality indicators not met
c	Three indicators not met
d	Four or more quality indicators not met

- b. definitions and criteria of all types of breathing events scored and used in comparisons.
6. Portable monitor methodology/definitions fully described: the polysomnography methods must include the following:
 - a. characterization of the equipment used;
 - b. definitions and criteria of all types of breathing events scored and used in comparisons
7. Portable monitor scoring fully described: includes a clear statement of whether manual or automated scoring was used, and, if automated, whether there was manual review/revision done.

Data Extraction

The strategy was taken to extract an amount of data that was succinct, complete, yet not overwhelming. The following data were extracted from the studies:

1. Device name/manufacturer
2. Author/year/reference number
3. Funding source/conflicts of interest
4. Test location (home vs. lab, lab vs. lab)
5. Attended/unattended
6. Hookup (technician or patient) and education offered
7. Timing of tests: simultaneous or interval between
8. Sample site
9. Inclusion/exclusion criteria
10. # Eligible/# Enrolled/# Completed (analyzed)
11. Demographics (% male, % non-white, mean age, mean BMI)
12. Comorbidities
13. Pretest probability or posttest prevalence
14. PSG flow and effort sensors
15. Scoring details
16. Results: ROC and AUC
17. Results: Sensitivity/specificity at AHI = 5
18. Results: Best Sensitivity (specificity) at what AHI
19. Results: Best (sensitivity) specificity at what AHI
20. Results: Other

APPENDIX II: EXCLUDED LITERATURE (TABLE A5)

Table A5—Excluded Literature

Author	SCOPER and Device	Reason
Abraham (2006) ³⁸	C ₁ O _{1x} E ₄ /ClearPath System (Nexan)	Wrong population (Patients with heart failure); screening for sleep apnea not OSA per se
Abraham (2008) ³⁹	C ₁ O _{1x} E ₄ /ClearPath System (Nexan)	Wrong population (Patients with heart failure); treatment of sleep apnea not OSA per se
Bagnato (2000) ⁴⁴	C ₄ O _{1x} R ₁ /Autoset	Therapeutic device used in diagnostic mode
BaHamman (2005) ⁴⁵	S ₁ C ₃ O _{1x} P ₂ E _x R ₃ (Alice 4)	Wrong outcomes measure (failure rate)
Berry (2008) ⁴⁶	S ₃ C ₂ O _{1x} /WatchPAT	Compared 2 clinical pathways, WatchPAT/APAP vs. PSG/CPAP; no data for LR+
Bridevaux (2007) ⁴⁷	C ₄ O _{1x} P ₂ E ₄ R ₂ (Embletta)	Wrong outcomes measure (ICC)
Chua (2007) ⁴⁸	S ₄ C ₃ O _{1x} P ₂ /Holter-oximeter (Northeast Monitoring)	Validation of sleep state measurement and derivation in patients with OSA
Chung (2010) ⁴⁹	S ₂ C ₁ O _{1x} P ₂ E ₁ R ₂ /Embletta X100	Manual vs. automated scoring
Freimark (2002) ⁴⁰	S ₃ C ₂ O _{1x} /WatchPAT	Wrong population (Patients with heart failure)
Galer (2007) ⁵⁰	O _{1x} R ₃ /SNAP	Comparison of snoring signal to RDI; no comparison to PSG
Gantner (2010) ⁵¹	O _{1x} and R ₂ vs. S ₂ C ₃ O _{1x} E ₁ R ₂ /ApneaLink vs. Somté	Only 1 channel of ApneaLink evaluated separately, nasal pressure or O ₂
Hedner (2004) ⁵²	S ₃ C ₂ O _{1x} /WatchPAT	Validation of sleep-wake algorithm only
Herscovici (2007) ⁵³	S ₃ C ₂ O _{1x} /WatchPAT	Validation of REM sleep algorithm only
Jurado-Gamez (2007) ⁵⁴	C ₄ O _{1x} P ₂ E ₁ R ₁ /Compumedics P-Series Screener	Only ICC and AUC reported
Levendowski (2009) ⁵⁵	S ₃ C ₄ O ₂ P ₂ E ₃ R ₂ /ARES Unicorder	Wrong outcomes measure (reliability)
Masdeu (2010) ⁵⁶	S ₃ C ₄ O ₂ P ₂ E ₃ R ₂ /ARES	Second analysis of Ayappa ¹⁴ data
Means (2004) ⁵⁷	SCOPER not adequately described	SCOPER not adequately described
Mueller (2006) ⁵⁸	C ₁ R ₃ /CardioMem	No oximeter
Mulgrew (2007) ⁵⁹	C ₄ O _{1x} E _x R _x /Remmers	Remmers used for pretest probability not diagnosis
Nakayama-Ashida (2008) ⁶⁰	S ₃ C ₄ O _{1x} P ₂ E ₁ R ₂ /Somté plus actigraphy	Wrong outcomes measure (ICC and night-to-night variability)
Penzel (2004) ⁶¹	S ₃ C ₂ O _{1x} /WatchPAT	Results cannot be converted to LR+
Pépin (2009) ⁶²	C ₁ R ₂ /SpiderView	No oximeter
Pillar (2003) ⁶³	S ₃ C ₂ O _{1x} /WatchPAT	Validation of arousal algorithm only
Popovic (2009) ⁶⁴	S ₃ C ₄ O ₂ P ₂ E ₃ R ₂ /ARES Unicorder	Wrong outcomes measure (effort sensor validation)
Prasad (2010) ⁶⁵	REMstar Pro M (CPAP device)	Only 1 channel (flow) and therapeutic device used in diagnostic mode
Quintana-Gallego (2004) ⁶⁶	S ₃ C ₃ O _{1x} P ₂ E _x R ₃ (Apnoescreen II)	Patients with heart failure
Skomro (2010) ⁶⁷	C ₄ O _{1x} P ₂ E ₁ R ₂ (Embletta)	ESS, PSQI, SF-36, BP, CPAP adherence
Smith (2007) ⁶⁸	O _{1x} P ₂ E ₄ R ₂ /Embletta	Patients with heart failure
Torre-Bouscoulet (2008) ⁶⁹	O _{1x} E _x R _x /Remmers/Snoresat	No outcome of interest
Whitelaw (2005) ⁷⁰	O _{1x} /Remmers/Snoresat	Only 1 channel
Yin (2005) ⁷¹	C ₄ O _{1x} P ₂ E ₂ R ₂ /Stardust II	Automated vs manual scoring

APPENDIX III: ADDITIONAL OUTCOMES INFORMATION ON DEVICES FROM EXCLUDED LITERATURE

Thermal Sensing Device

In a Level IVa study, BaHamman⁴⁷ [$S_1C_3O_{1x}P_2E_xR_3$ (Alice 4)] reported on the failure rate of signals and sensors when using the same equipment attended in-lab versus unattended in a hospital. Hook-up was by a trained technologist and there was a pre-investigation into causes of failures of signals and modifications thereof. The failure rate for the different signals ranged from 0.128 min in electrocardiography (EKG) to 67.36 min in the thoracic belt signal. However, that did not affect the success rate of the studies. Acceptable scorable data was available in 97% of the performed unattended PSGs.

Quintana-Gallego et al.⁶⁸ [$S_3C_3O_{1x}P_2E_4R_3$ (Apnoescreen II, effort not defined, but assumed to be the same as that reported for Apnoescreen II from Garcia-Diaz et al.⁹)] reported on a cohort of patients with stable congestive heart failure (CHF). Since this is a specific population with comorbidity, these data were not included in this evaluation. The paper was Level Ib, was conducted in the home and lab setting, and good LR+ results were obtained (7.2 at $AHI \geq 5$ and 39.5 at $AHI \geq 10$) on 90 enrolled patients (75 completed).

Nasal Pressure Devices

Bridevaux et al.⁴⁹ reported in a Level IVb study on the interobserver agreement for a home study of a $C_4O_{1x}P_2E_4R_2$ (Embletta) device (that used 2 piezoelectric effort belts). The ICC for AHI was reported to be 0.73.

Skomro et al.⁶⁹ reported in a Level Ib study on the use of a $C_4O_{1x}P_2E_1R_2$ (Embletta) device using home vs. lab management schemes. They reported that there were no significant differences in ESS, PSQI, SF-36, BP, or CPAP adherence when the subjects were diagnosed and prescribed CPAP treatment with the OOL device vs in-lab PSG.

One study each reported on reliability [Levendowski et al.⁵⁷ Level IVb home/lab reliability] and effort sensor validation [Popovic et al.⁶⁶ Level IIa Lab/Lab data on effort sensors] of the ARES Unicorder ($S_3C_4O_2P_2E_3R_2$). Levendowski et al. reported that night-to-night variability in the home was actually 50% less than that in the lab. Popovic et al. reported in a small study (n = 14 completed) on the intrarater and interrater reliability of forehead venous pressure as a measure of effort as an alternative to esophageal manometry, chest, or abdominal piezoelectric belts for the measurement of respiratory effort. With respect to interrater κ scores, the chest belt was superior to the other measures. For intrarater agreement versus the gold standard of esophageal manometry, the other 3 measures

showed near perfect agreement. FVP was superior to either effort belt in the detection of obstructive apneas and hypopneas, similar in the detection of persistent flow limitation and physiological changes in ventilation, and inferior in the detection of central events.

Nakayama-Ashida et al.⁶² reported on the reliability of the Somté plus actigraphy ($S_3C_4O_{1x}P_2E_1R_2$) in a Level IIIb home study. The ICC was 0.98 for interscorer reliability and 0.95 for night-to-night variability.

Chung et al.⁵¹ reported in a Level Ia lab/lab study on the validity of the Embletta X100 $S_2C_1O_{1x}P_2E_1R_2$ versus in-lab attended PSG that manual scoring was superior to automated scoring for reliability. In a second part of the study (Level IIIa), in a perioperative home or hospital unattended setting, 88.7% of the recordings were technically good, 9% technically acceptable, and 2.3% were failures.

Yin et al.⁷³ used $C_4O_{1x}P_2E_2R_2$ Stardust II in a Level IIIa home study of automated vs. manual scoring. At $AHI < 30$, the maximum LR+ was 2.6; for $AHI \geq 40$ the LR+ was 5.5, and for $AHI \geq 50$, the LR+ was 13.3. The authors concluded that analyses should be done manually.

Smith et al.⁷⁰ studied a population of patients with chronic heart failure with $O_{1x}P_2E_4R_2$ (Embletta) in a Level Ib study with both H/L and L/L comparisons on 20 patients. Two piezoelectric belts were used. LR+ was calculated as 2 (not clinically useful). Scoring depended on either a reduction in airflow or thoracic-abdominal movements. Since piezoelectric belts were used, this could explain the poor agreement with lab PSG. Alternative conclusions are that this population should not be studied out-of-lab or that nasal pressure without a thermal sensor is not recommended in this situation.

Thermal Sensing Device Plus Nasal Pressure Devices

Jurado-Gómez et al.⁵⁶ reported in a Level Ib home/lab study on the performance of the Compumedics P-Series Screener ($C_4O_{1x}P_2E_1R_1$). Fifty-two patients completed the study. No ROC curves were presented, but the AUCs were reported as 0.804 at $AHI \geq 10$ and 1.00 at $AHI \geq 30$. The diagnoses coincided in 88.4% of cases with an ICC of 0.963. Using a visual analog scale, patients preferred home testing over lab PSG testing, with a median score of 7 for lab PSG and 9 for home testing ($p < 0.0001$).

PAT Devices ($S_3C_2O_{1x}$ /WatchPAT)

A Level IVa by Berry et al.⁴⁸ compared 2 clinical pathways, WatchPAT/APAP vs. PSG/CPAP. They reported no statistically significant differences in diagnostic testing satisfaction, titration satisfaction, CPAP adherence, number of patients stopping before 6 weeks, average nightly use, or outcomes [including ESS, FOSQ, CPAP satisfaction, or machine estimate of residual AHI] between home/APAP and lab/CPAP management of patients.