

Technology Assessment



**Technology
Assessment Program**

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

**Home diagnosis of Obstructive
Sleep Apnea-Hypopnea Syndrome**

August 8, 2007

Home Diagnosis of Obstructive Sleep Apnea-Hypopnea Syndrome

Thomas A. Trikalinos, MD, PhD
Stanley Ip, MD
Gowri Raman, MD
M. Soledad Cepeda, MD, PhD
Ethan M. Balk, MD, MPH
Carolyn D'Ambrosio, MS, MD
Joseph Lau, MD

August 8, 2007

Tufts-New England Medical Center EPC

Table of contents

Table of contents.....	5
Table of contents for Figures.....	9
Table of contents for Tables.....	11
Summary.....	12
Introduction.....	15
Obstructive Sleep Apneas-Hypopnea Syndrome (OSAHS).....	15
Assessing the presence and quantifying the severity of OSAHS.....	15
Statement of Work.....	16
Specific aims.....	16
Follow-on project.....	17
Methods.....	19
Terminology and definitions.....	19
Disease.....	19
Respiratory events.....	19
Sleep time.....	20
Apnea-Hypopnea Index (AHI) and Respiratory Disturbance Index (RDI).....	20
Cardiorespiratory and neurophysiologic signals.....	21
Types of monitors used in sleep studies.....	22
Other definitions and terms.....	22
Search Strategy.....	23
Inclusion and exclusion criteria.....	24
Participants.....	24
Interventions and assessments.....	24
Eligible outcomes - eligible methods of analyses.....	25
Design.....	26
Data abstraction.....	26
Analyses.....	27
Assessing whether baseline AHI or RDI predicts response to CPAP.....	27
Comparison of measurements with different methods.....	27
Comparison of rates of data loss, errors, harms, adverse events and complications.....	28
Assessing individual studies.....	28
Applicability and quality of studies.....	28
Results.....	31
Specific Aim A: Clarifications of important topics.....	31
A1. Is laboratory-based polysomnography considered a “gold standard” for the diagnosis of sleep apnea?.....	31
Does laboratory-based PSG measure AHI reliably?.....	31
Is AHI (or RDI) sufficient to diagnose OSAHS?.....	33
Is AHI necessary for the management of people suspected for OSAHS?.....	34
How should portable monitors be evaluated?.....	35
A2. Appropriate methods for the comparison of diagnostic test performance for obstructive sleep apnea.....	35
Study design issues.....	35

Issues on statistical analyses in the comparison of measurements with two methods.....	36
A3. Classification of portable sleeping monitoring devices.....	42
A3a. Different parameters of sleep and diagnosis of OSAHS.....	42
Specific Aim B: Role of portable monitors vs. facility-based polysomnography in the diagnosis of OSAHS.....	45
Literature flow.....	45
B1. Ability of facility-based polysomnography vs. portable monitors to predict response to CPAP and changes in clinical outcomes after CPAP treatment.....	46
Associations of indices from sleep studies with adherence to CPAP use.....	46
Associations of indices from sleep studies with response to CPAP in terms of quality of life outcomes.....	48
Associations of indices from sleep studies with response to CPAP in terms of other outcomes.....	50
Synopsis for section B1.....	51
B2. How does the performance of portable monitors compare with facility-based polysomnography for the diagnosis of OSAHS?.....	52
Type II monitors.....	52
Type III monitors.....	56
Type IV monitors.....	74
Synopsis for section B2.....	98
B2a. For studies in the home setting, do technologist support and patient education affect the comparison of portable monitors and with facility-based polysomnography for the diagnosis of OSAHS?.....	101
Type II monitors.....	101
Type III monitors.....	101
Type IV monitors.....	101
Synopsis for section B2a.....	102
B3a. How do automated and manual scoring compare in the diagnosis of obstructive sleep apnea?.....	102
Facility-based polysomnography (Type I monitors).....	102
Type II monitors.....	102
Type III monitors.....	102
Type IV monitors.....	105
Synopsis for section B3a.....	106
B3b. What errors related to automated and manual scoring are reported?.....	106
Synopsis for section B3b.....	106
B4. For studies of portable monitoring in the home setting, what errors related to unattended use are reported?.....	106
Type II monitors.....	106
Type III monitors.....	107
Type IV monitors.....	107
Synopsis for section B4.....	108
B5. Do the reported complications, harms, and adverse events differ for portable multichannel sleep testing and facility-based polysomnography?.....	108
FDA data on adverse events of medical devices used in sleep studies.....	108

Laboratory-based polysomnography (Type I monitors).....	108
Type II monitors	109
Type III monitors	109
Type IV monitors.....	109
Synopsis for section B5	109
B6. Do rates or types of data loss and data corruption differ for portable multichannel sleep testing and facility-based polysomnography?.....	109
Synopsis for section B6	112
Overview.....	113
Key Question A1.....	113
Key Question A2.....	113
Key Question A3.....	114
Key Question B1.....	114
Key Question B2.....	114
Key Question B3.....	117
Key Question B4.....	117
Key Question B5.....	117
Key Question B6.....	117
References and included studies	119

Table of contents for Figures

Figure 1. Total sleep time and total recording time.....	20
Figure 2. Illustrative example: measurements with facility-based PSG and type IV monitors in Ayappa 2004 (digitized data).	37
Figure 3. Illustrative example: difference versus average analyses of measurements with facility-based PSG and type IV monitors in Ayappa 2004 (digitized data).	39
Figure 4. Square plot of sensitivity versus 100%-specificity.	41
Figure 5. Literature flow	45
Figure 6. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type II monitors.	55
Figure 7. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type III monitors in the home setting.	61
Figure 8. Diagnostic ability of type III monitors in the home setting to identify AHI suggestive of OSAHS in laboratory-based polysomnography.	62
Figure 9. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type III monitors in the hospital setting (not specialized sleep units).	63
Figure 10. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type III monitors in specialized sleep units. Studies that used manual scoring for the portable monitor.	68
Figure 11. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type III monitors in specialized sleep units. Good and moderate quality studies that used automated scoring for the portable monitor.	69
Figure 12. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type III monitors in specialized sleep units. Poor quality studies that used automated scoring for the portable monitor.	70
Figure 13. Diagnostic ability of type III monitors in specialized sleep units to identify AHI>15 events/hour in laboratory-based polysomnography. Manual or combined manual and automated scoring.....	71
Figure 14. Diagnostic ability of type III monitors in specialized sleep units to identify AHI>15 events/hour in laboratory-based polysomnography. Automated scoring. ...	72
Figure 15. Diagnostic ability of type III monitors in specialized sleep units to identify AHI>10 events/hour in laboratory-based polysomnography. Manual or combined manual and automated scoring.....	73
Figure 16. Diagnostic ability of type III monitors in specialized sleep units to identify AHI>10 events/hour in laboratory-based polysomnography. Automated scoring. ...	73
Figure 17. Diagnostic ability of type III monitors in specialized sleep units to identify AHI>20 events/hour in laboratory-based polysomnography.....	74
Figure 18. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type IV monitors (with at least 3 channels) in specialized sleep units.....	80
Figure 19. Diagnostic ability of type IV monitors in the home setting to identify AHI>15 events/hour in laboratory-based polysomnography.	82

Figure 20. Diagnostic ability of type IV monitors in the home setting to identify AHI>10 events/hour in laboratory-based polysomnography.	82
Figure 21. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type IV monitors (with at least 3 versus 1 or 2 channels) in specialized sleep units. Studies with good or moderate overall methodological quality.	92
Figure 22. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type IV monitors (with at least 3 versus 1 or 2 channels) in specialized sleep units. Studies with poor overall methodological quality.	93
Figure 23. Diagnostic ability of type IV monitors (with at least 3 versus 1 or 2 channels) in specialized sleep centers to identify AHI>15 events/hour in laboratory-based polysomnography. Studies using manual scoring for the type IV monitor.	95
Figure 24. Diagnostic ability of type IV monitors (with at least 3 versus 1 or 2 channels) in specialized sleep centers to identify AHI>15 events/hour in laboratory-based polysomnography. Studies using automated scoring for the type IV monitor.	96
Figure 25. Diagnostic ability of type IV monitors (with at least 3 versus 1 or 2 channels) in specialized sleep centers to identify AHI>10 events/hour in laboratory-based polysomnography. Studies using manual scoring for the type IV monitor.	96
Figure 26. Diagnostic ability of type IV monitors (with at least 3 versus 1 or 2 channels) in specialized sleep centers to identify AHI>10 events/hour in laboratory-based polysomnography. Studies using automated scoring for the type IV monitor.	97
Figure 27. Diagnostic ability of type IV monitors (with at least 3 versus 1 or 2 channels) in specialized sleep centers to identify AHI>20 events/hour in laboratory-based polysomnography.	97
Figure 28. Diagnostic ability of type III monitors in different settings to identify AHI>15, >10 and >20 events/hour in laboratory-based polysomnography. Comparison of scoring methods.	104
Figure 29. Diagnostic ability of type IV monitors in different settings to identify AHI>15 events/hour in laboratory-based polysomnography. Comparison of scoring methods.	105
Figure 30. Rates of unsatisfactory sleep recordings for facility-based PSG and portable monitors in studies that clearly reported the pertinent information.	111
Figure 31. Comparative rates of unsatisfactory sleep recordings for facility-based polysomnography and portable monitors.	111

Table of contents for Tables

Table 1. Often used signals in sleep studies.	21
Table 2. Delineation of operational rules used to classify monitors in sleep studies.	22
Table 3. 2 by 2 table used in the calculation of sensitivity and specificity of portable monitors to predict AHI suggestive of OSAHS in facility-based PSG.	40
Table 4. Associations of baseline apnea-hypopnea index with CPAP use.	47
Table 5. Associations of baseline apnea-hypopnea index with response to CPAP in terms of quality of life outcomes.	49
Table 6. Associations of baseline apnea-hypopnea index with outcomes other than those assessed in Tables 4 and 5.	51
Table 7: Description of studies comparing type II monitors with facility-based PSG.	53
Table 8. Sensitivity and specificity of type II monitors in the sleep laboratory setting to predict AHI>15 events/hour with laboratory-based PSG.	56
Table 9. Description of studies comparing type III monitors with facility-based PSG in settings other than a specialized sleep clinic or unit.	58
Table 10. Description of studies comparing type III monitors with facility-based polysomnography in the sleep clinic or sleep unit.	65
Table 11: Description of studies comparing type IV monitors (three or more bioparameters) with facility-based polysomnography in settings other than a specialized clinic or unit.	76
Table 12: Description of studies comparing type IV monitors (one or two bioparameters) with facility-based polysomnography in the home setting.	78
Table 13. Description of studies comparing type IV monitors (recording three or more bioparameters) with facility-based polysomnography in specialized sleep units.	84
Table 14. Description of studies comparing type IV monitors (recording one or two bioparameters) with facility-based polysomnography in specialized sleep units.	87
Table 15. Significant predictors of valid signal duration in unattended home monitors (type II) in the Sleep Heart Health Study.	107

Summary

Obstructive Sleep Apnea-Hypopnea Syndrome (OSAHS), or Obstructive Sleep Apnea -OSA, is characterized by sleep disturbances secondary to upper airway obstruction. OSAHS is prevalent in two to four percent of middle-aged adults, and has been associated with daytime somnolence, cardiovascular morbidity, diabetes and other metabolic abnormalities, and increased likelihood of accidents and other adverse outcomes. The prevalence of OSAHS in older adults (65 years or older) is believed to be higher than the aforementioned estimates, but it is not as well studied. In fact, there are contradictory data suggesting that OSAHS prevalence levels off after the age of 65 years; a plausible explanation is that fewer new cases are identified among older adults. Continuous positive airway pressure (CPAP) is the commonly used treatment (additional interventions exist, but are not used or indicated in the majority of patients).

The reference standard for the diagnosis of OSAHS is facility-based polysomnography (PSG), a comprehensive sleep study that records and evaluates a variety of cardiorespiratory and neurophysiologic signals during sleep time. It quantifies the severity of disturbances with the Apnea-Hypopnea Index (AHI). Higher AHI values imply more severe sleep disturbances. Typically, a value of 15 events/hour of sleep or more is considered to be *suggestive of OSAHS*. Note that an AHI suggestive of OSAHS is not sufficient for the *diagnosis* of the condition, as the severity of symptoms has to be accounted for, and other conditions affecting sleep may need to be excluded.

Portable monitors have been developed in an effort to substitute for the more costly facility-based PSG. According to the data recorded, monitors are classified into different categories. Facility-based PSG is a type I monitor. Type II portable monitors record the same data as facility-based PSG, albeit using fewer channels. Type III portable monitors are less comprehensive, since they do not record neurophysiologic sleep staging information. Type IV portable monitors are those that fail to meet criteria for type III devices (i.e., they do not record at least two respiratory channels).

This technology assessment is based on a systematic review of the literature. In total 95 studies were included. Eligible studies assess the ability of sleep studies at baseline to predict response to CPAP treatment or CPAP use; the comparison of measurements with portable monitors and facility-based PSG; and the safety of sleep studies. We note that the current technology assessment focuses on adult patients, and did not consider pediatric populations.

Baseline AHI is only modestly associated with response to CPAP or CPAP use among people with high (pre-test) probability for OSAHS. The same is true for other indices obtained from sleep studies such as the mean or minimum O₂ saturation, apnea index, hypopneas index, frequency of arousals and other quantities. Note that none of the eligible studies assessed hard clinical outcomes (i.e., mortality, myocardial infarctions, strokes and similar outcomes).

Based on limited data, type II monitors may identify AHI suggestive of OSAHS with high positive likelihood ratios (>10) and low negative likelihood ratios (<0.1) both when the portable monitors were studied in the sleep laboratory and at home. Type III monitors may have the ability to predict AHI suggestive of OSAHS with high positive likelihood ratios and low negative likelihood ratios for various AHI cutoffs in laboratory-based PSG, especially when manual scoring is used. The ability of type III monitors to predict

AHI suggestive of OSAHS appears to be better in studies conducted in the specialized sleep unit compared to studies in the home setting. Studies of type IV monitors that record at least three bioparameters showed high positive likelihood ratios and low negative likelihood ratios. Studies of type IV monitors that record one or two bioparameters also had high positive likelihood ratios and low negative likelihood ratios, at least for selected sensitivity and specificity pairs from ROC curve analyses.

Similarly to type III monitors, the ability of type IV monitors to predict AHI suggestive of OSAHS appears to be better in studies conducted in specialized sleep units. However, Medicare beneficiaries are older than the studied subjects (the median average age was approximately 50 years in the analyzed studies), and may have more often comorbidities that affect sleep (i.e., non-OSAHS conditions such as cardiac insufficiency; chronic obstructive pulmonary disease; obesity hypoventilation syndrome; or Periodic Limb Movements in Sleep and Restless Leg Syndrome). These conditions may be misdiagnosed as OSAHS by sleep monitors that do not record channels necessary for the differential diagnosis from OSAHS. Therefore, some type III and type IV monitors may yield more false positives among Medicare beneficiaries, compared to what was observed in the assessed studies. We stress that in the assessed studies the frequency of patients with the aforementioned comorbidities was very low (in some these patients were clearly excluded, in the remaining the frequency was very low or not reported at all).

For studies in the home setting, there is no direct data on whether and to what extent technologist support and patient education affect the comparison of portable monitors with facility-based PSG.

Overall, manual scoring or manual editing of automated scoring seems to have better agreement with facility-based PSG compared to manual scoring in the studies that assessed both scoring methods. We note that the automated scoring algorithms may vary across different monitors, or even with the specific software version or settings. Thus their ability to recognize respiratory events may differ.

We did not identify detailed data on the specific types of errors that are directly related to automated or manual scoring. No studies associated directly any specific errors with unattended use. However, signal loss was more often observed in home studies, and one study associated discrepancies in the AHI measurement with poor quality airflow signals in the unattended home-based recordings.

The rate and severity of adverse events in sleep studies is low. In a large study of over 16,000 facility-based PSG complications were identified in less than 0.5% of the recordings. Complications were not reported in the remaining studies, and mostly minor harms were reported to the FDA adverse events database. This conclusion applies to both facility-based PSG and to portable monitors in various settings, including the home setting.

Rates of unsatisfactory studies and data corruption are higher for portable monitors in the home setting, compared to facility-based PSG, or portable monitors in the sleep laboratory setting. This may be attributed to user errors during device operation or probe hook-up for home studies, and also to the absence of an attending technologist.

Introduction

Obstructive Sleep Apneas-Hypopnea Syndrome (OSAHS)

Obstructive Sleep Apnea (OSA) is a relatively common disorder that could affect all ages. The condition is characterized by periods of disturbed airflow patterns during sleep time, namely reduced airflow (hypopnea) or airflow cessation (apnea). It is postulated that both types of airflow disturbance have similar pathophysiology and bear the same clinical significance.¹ Obstructive sleep apnea-hypopnea syndrome (OSAHS) is the most common type of the condition (apneas and hypopneas of central and mixed central and obstructive etiology comprise the other forms).¹ OSAHS has been associated with a variety of adverse clinical outcomes such as mortality secondary to stroke and cardiovascular events,^{2,3} decreased quality of life,⁴ cardiovascular disease and stroke,^{2,5} hypertension,⁶⁻⁸ diabetes and other metabolic abnormalities,^{9,10} as well as increased likelihood for driving^{11,12} and other types of accidents.

Assessing the presence and quantifying the severity of OSAHS

The severity of OSAHS is typically quantified by the number of apneas and hypopneas per hour of sleep, a quantity that has been termed Apnea-Hypopnea Index (AHI). Different populations have different AHI values. Specific cutoffs are typically used to establish the diagnosis of OSAHS.^{1,13} For example, as of this writing, the Medicare criteria for reimbursement of continuous positive airway pressure (CPAP) therapy are AHI ≥ 15 events/hour, or AHI ≥ 5 events/hour associated with symptoms (e.g., daytime somnolence and fatigue). However, a variety of AHI thresholds ranging between 5 and 40 have been used as suggestive of OSAHS in different studies.

Approximately two to four percent of middle-aged women and men, respectively, have been estimated to have an AHI ≥ 15 events/hour and excessive daytime somnolence in the population-based Wisconsin Sleep Cohort Study.⁴ Using an AHI cutoff of ≥ 5 events/hour without the symptoms of excessive daytime sleepiness puts the prevalence at 9% for women and 24% for men. The symptom of excessive daytime sleepiness is quite variable and not always present in patients with OSAHS. Thus, most people suffering from OSAHS remain undiagnosed and untreated.⁴ More recent studies also suggest a high prevalence (i.e., prevalence of AHI ≥ 5 in adults age 30-69 of 17%), perhaps due to increasing obesity rates in later years.¹⁴

The prevalence of the condition among Medicare beneficiaries (people aged 65 years or older) is believed to be higher than the aforementioned estimates among middle-aged people. In the population-based Sleep Heart Health Study the prevalence of AHI ≥ 15 events/hour was 1.7-fold higher in people older than 60 years compared to people between 40 and 60 years of age.¹⁵ Similar observations were made in cohort studies that used population-based samples and a wide range of ages.¹⁶⁻¹⁹ However, scant data suggest that the prevalence of OSAHS does not continue to rise with age in older adults, but reaches a plateau after the age of 60-65 years.^{15,20} This implies either a relative increase in mortality from OSAHS, or a remission of OSAHS with advancing age.

Apart from the use of AHI, other methods to quantify severity have also been used in various studies. These mainly pertain to the evaluation of O₂ desaturations during sleep, the evaluation of other respiratory events such as the Respiratory Effort Related Arousals, or the degree of daytime fatigue and somnolence.

The standard measurement of AHI (and the diagnosis of OSAHS by extension) requires a comprehensive, technologist-attended sleep study with multichannel polysomnography (PSG), which is performed in specialized sleep laboratories.^{1,13} Laboratory-based PSG records a variety of neurophysiologic and cardiorespiratory signals and is interpreted by trained technologists and sleep physicians after the sleep study has been completed. Because of the high demand, the associated costs and the need for timely diagnosis, portable devices have been developed to substitute for laboratory-based PSG.¹³ There are different types (classes) of portable monitors.²¹ Each gathers different neurophysiologic and respiratory information and may synthesize the accumulated data differently.²¹ Depending on the data they record, portable monitors are classified in different categories (which are discussed in more detail later in this technology assessment).²¹ Portable monitors can be used not only in the home setting, but in the hospital and clinics other than specialized sleep units.

Statement of Work

The Center for Medicare and Medicaid Services (CMS) has requested a technology assessment through the Agency for Healthcare Research and Quality (AHRQ) on the role of home monitoring for the diagnosis of OSAHS. On September 28, 2004, the evidence on home monitoring devices in the diagnosis of sleep apnea was discussed at a Medicare Coverage Advisory Committee meeting. The RTI EPC presented a technology assessment on this topic,²² which was an update of a prior technology assessment done for the American Academy of Sleep Medicine, the American Thoracic Society, and the American College of Chest Physicians.¹³ CMS has requested an update of the evidence presented in the RTI EPC technology assessment on home sleep monitoring with an expanded scope, including the assessment of the ability of PSG indices to predict a response to CPAP treatment. More specifically, the following aims were defined by CMS and AHRQ after discussions with the Tufts-NEMC EPC:

Specific aims

- A. Provide a discussion on the following topics:
 - A1. Discuss whether facility-based polysomnography is considered a “gold standard” for the diagnosis of sleep apnea.
 - A2. Discuss the appropriate methods for the comparison of diagnostic test performance for obstructive sleep apnea.
 - A3. Discuss the classification of sleep monitoring devices (Types I, II, III and IV).
 - A3a. Discuss the relationship between the measurement of parameters of sleep and cardiorespiratory function (i.e., sleep staging, body position, limb movements, respiratory effort, airflow, oxygen saturation, electrocardiogram [ECG]), and the accuracy of diagnosis of sleep apnea.
- B. Summarize the scientific evidence on portable monitoring for the diagnosis of obstructive sleep apnea. In particular, summarize evidence on the following questions:
 - B1. Among patients diagnosed with obstructive sleep apnea (OSA) and considered for treatment with CPAP (of any form), how does baseline diagnosis by portable

- polysomnography compare with facility-based polysomnography in predicting the response to CPAP and clinical outcomes?
- B2. How does the performance of portable multi-channel sleep testing compare with facility-based polysomnography (full or split night) for the diagnosis of obstructive sleep apnea?
 - B2a. For studies of portable monitoring in the home setting, how do factors such as technologist support and patient education affect the comparison of portable multi-channel sleep testing with facility-based polysomnography for the diagnosis of obstructive sleep apnea?
 - B3a. How do automated and manual scoring compare in the diagnosis of obstructive sleep apnea?
 - B3b. What errors related to automated and manual scoring are reported?
 - B4. For studies of portable monitoring in the home setting, what errors related to unattended use are reported?
 - B5. Do the reported complications, harms, and adverse events differ for portable multichannel sleep testing and facility-based polysomnography?
 - B6. Do rates or types of data loss and data corruption differ for portable multichannel sleep testing and facility-based polysomnography?

Follow-on project

Finally, as a follow-on project, Tufts-New England Medical Center EPC will create a decision model to assess management strategies for patients with high clinical suspicion for OSAHS.

Methods

This technology assessment has two aims, as described in the Statement of Work. Aim A is a discussion of important relevant topics and is based on narrative reviews. Aim B is based on a systematic review of the literature.

Terminology and definitions

For this report, we will use the following definitions and terms throughout the technology assessment.

Disease

Obstructive Sleep Apnea Hypopnea Syndrome and Obstructive Sleep Apnea are terms that refer to the same entity. The latter name is used in the International Classification of Diseases (ICD) scheme and in the standard classification in the International Classification of Sleep Disorders (ICSD) – Diagnostic and Coding Manual. However, we retain the more descriptive OSAHS throughout this technology assessment.

Respiratory events

OSAHS is characterized by the presence of respiratory events during sleep. The following respiratory events are of interest (all events are defined for a minimum duration of 10 seconds):

- *Apnea* is the cessation of airflow. Some studies have defined apnea as a very large reduction in airflow (more than approximately 80% of baseline), but not necessarily complete cessation; this is because of the inability of airflow probes to differentiate complete cessation from very low flow.
- *Hypopnea* is a reduction in airflow compared to “baseline” airflow. The definition of “baseline” airflow may influence the number of hypopneas that are scored, especially in patients with lung disease and fluctuating ventilation. The definition of “baseline” airflow is not reported in most papers. Definitions for hypopnea vary across studies. Hypopneas have been defined as reductions in airflow alone or reductions in airflow associated with other events (such as arousals or oxygen [O₂] desaturations). The magnitude of airflow reduction may be quantitatively or qualitatively defined. Usually employed quantitative definitions identify hypopneas with reductions in the amplitude of the airflow signal of a given magnitude (e.g., at least 50%). Studies employing qualitative definitions identify hypopneas with “discernible airflow reductions” (again due to the inability of airflow probes to reliably quantify low flow).
- *Other respiratory events*. Depending on the signals gathered during the sleep study, definitions of the respiratory events may differ. For example, if only oximetry were used, the degree and frequency of desaturations would be assessed as proxies for apneic and hypopneic respiratory events. Additionally, recent evidence suggests that arousals from sleep due to respiratory efforts against a blocked or partially blocked upper airway contribute significantly to cardiovascular disease in these patients. An additional respiratory event that has been described is the Respiratory Effort Related Arousal (RERA). This occurs when there is a discernible reduction in the nasal pressure cannula

signal associated with an arousal from sleep. RERAs are validated by esophageal pressure monitoring.^{1,23,24}

Sleep time

- *Total recording time* or *total time in bed* is the time period a person spent in bed or the total duration of the sleep study. This time interval is otherwise referred to as the time that lapsed between lights off and lights on. It is measured in facility-based PSG to determine overall *sleep efficiency* (total sleep time divided by total recording time), and it is measured in portable monitor studies as well. In many portable monitor studies it is used as a proxy for the actual total sleep time.
- *Total sleep time* pertains to the total time a person spent sleeping, and is by definition shorter than the *total recording time*. To estimate the total sleep time, sleep staging must be performed. The most often employed criteria for sleep staging in sleep studies are the criteria of Rechtschaffen and Kales.²⁵

Figure 1 illustrates the distinction between total recording time and total sleep time.

Figure 1. Total sleep time and total recording time.

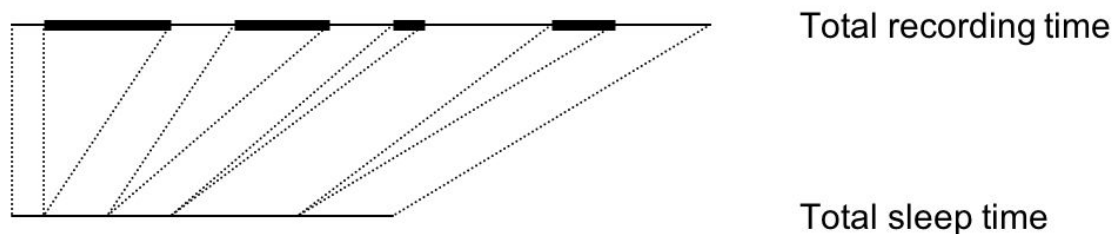


Figure 1. The total recording time in a sleep study consists of the total time spent asleep (horizontal thin line segments) and the time the patient is aroused (horizontal bold line segments). The total sleep time excludes arousal periods and is by definition shorter or at most (i.e., in the absence of arousals) as long as the total recording time. (N.B., This figure is only a schematic and the lengths of the arousal or sleep periods or their numbers are not drawn to scale).

Apnea-Hypopnea Index (AHI) and Respiratory Disturbance Index (RDI)

The frequency of respiratory events during sleep can be used to quantify the severity of OSAHS.

$$\text{frequency of respiratory events} = \frac{N_{\text{respiratory events}}}{\text{Total sleep time}}$$

The following two indices are used to quantify the severity of OSAHS in most studies. Other indices have also been used.

- *Apnea-Hypopnea Index (AHI)*. The number of apneas and hypopneas per hour of sleep. It is a fraction, where the numerator is the number of apneas and hypopneas, and the denominator is total sleep time.
- *Respiratory Disturbance Index (RDI)*. We caution that studies of portable monitors do not define RDI in the same way it is defined in everyday clinical practice in the sleep laboratory (i.e., the quotient of the total number of RERAs, apneas and hypopneas divided by total sleep time). In studies of portable monitors RDI is a quantity that

approximates AHI, whenever the numerator (apneas or hypopneas) or the denominator (total sleep time) or both are not measured directly. In most cases the denominator is the total recording time instead of the total sleep time. Proxies for the numerator vary depending on the recorded signals and their assessment. Therefore exact definition of RDI may vary across different studies of portable monitors. Throughout the technology assessment, we refer to RDI in the way it is defined in portable monitor studies.

- Other indices are not used as extensively or may be signal-specific. The *apnea index (AI)* is defined similarly to the AHI, but ignoring hypopneas in the numerator (e.g., see Fleury 1996²⁶). In some studies of oximetry-only recordings, the oxygen desaturation index (ODI) is used. This is typically the number of times the O₂ saturation drops by a certain percentage (4% usually) per hour of recording time. Other oximetry studies measure the “delta index” (it assesses serial changes in O₂ saturation, e.g., Pepin 1991²⁷). Snoring intensity (e.g., Esnaola 1996 and Koziej 1994,^{28,29}) or other indices have also been used.

Cardiorespiratory and neurophysiologic signals

During sleep studies, several types of data (signals) are recorded in the pertinent monitor channels and are then evaluated. These signals are physiologic measurements. **Table 1** provides a simplified description and explanation of the use of different signals that are usually recorded in sleep studies. This list is not exhaustive.

Table 1. Often used signals in sleep studies.

Signal	Description	Probes*
<i>Signals used to define respiratory events (“numerator”)</i>		
Airflow	To define apneas or hypopneas	Thermistor (nasal, oronasal); nasal cannula; tracheal microphone
Effort	Respiratory effort, movement of the diaphragm or the rib cage	Thoracoabdominal piezoelectric belts; inductance plethysmography; intercostal-muscle EMG; diaphragm EMG
SaO ₂	O ₂ saturation of arterial blood	Oximetry (ear lobe; digit)
<i>Signals used to define sleep time (“denominator”)</i>		
EOG	To identify REM sleep	Electrodes (may be bilateral)
EEG	Cerebral activity	Electrodes (e.g., 4 electrodes)
Chin (submental) EMG	Increased tone is associated with arousal	Electrodes (may be bilateral)
Anterior tibialis EMG	Increased tone is associated with arousal	Electrodes (may be bilateral)
ECG	To measure heart rate or to quantify its variability (an estimate of sympathetic activity)	Usually a single electrode to estimate the RR interval
Body position sensor	Describes sleep position	e.g., Mercury switch
Video monitoring	Describes sleep position, differentiate periods of mouth breathing from apneas, etc.	Infrared-sensitive camera
Actigraphy sensor	Describes acceleration and movement of body parts	e.g., Wrist actigraphy

* the list is not exhaustive

ECG: electrocardiogram; EEG: Electroencephalogram; EMG: electromyography; EOG: Electro-oculogram; REM: Rapid eye movement

Note that the list of channels is not exhaustive. Refer to text for a description of channels used by different types of monitors.

Types of monitors used in sleep studies

The American Sleep Disorders Association classified the different monitors that have been used in sleep studies into four categories, depending on which channels they record and evaluate.²¹ Details on the classification of sleep monitors and a discussion on how newer devices may fit in this classification scheme are provided in Section A3.

Here, we used the operational rules described in **Table 2** to classify sleep monitors. Very similar rules have been applied in previous systematic reviews.¹³ Briefly:

- Type I is facility-based PSG.
- Type II monitors record the same information as type I (perhaps with fewer channels). Type II monitors record signals that allow the reliable identification of (micro)arousals from sleep (e.g. EOG, chin EMG, EEG – see **Table 1** for abbreviations) *and* at least two respiratory channels (two airflow channels or one airflow and one effort channel).
- Type III monitors do not record the channels that differentiate between sleep and wake, but have at least two respiratory channels (two airflow channels or one airflow and one effort channel).
- Type IV are all other monitors that fail to fulfill criteria for type III monitors. Therefore type IV channels may include monitors that record more than 2 bioparameters.

We classified each monitor according to the channels *that were actually used* in the pertinent study. For example, if not all channels of a nominally type III monitor were used or analyzed, we classified the monitor as “type IV” for the particular study. This “downlabeling” occurred rarely; it is always clearly noted in the corresponding tables and text.

Table 2. Delineation of operational rules used to classify monitors in sleep studies.

Type or Level	Portability	Indicative N _{channels}	Indicative signals	≥2 airflow /effort channels	Identifies sleep /wake	AHI
I	Facility-based	~14-16	EEG, EOG, EMG, ECG/HR, airflow, effort, SaO ₂	Yes	Yes	Yes
II	Portable	≥7	(may have EEG), HR*, EOG, chin EMG, ECG/HR, airflow, effort, SaO ₂	Yes	Yes	Yes
III	Portable	≥4	Airflow and/or effort, ECG/HR, SaO ₂	Yes	No	No
IV	Portable	~1-3**	[All monitors not qualifying for type III]	No	No***	No

AHI: Apnea-Hypopnea index; ECG: electrocardiogram; EEG: Electroencephalogram; EMG: electromyography; EOG: Electro-oculogram; HR: heart rate; SaO₂: arterial O₂ saturation;

* Heart rate is allowed instead of EEG in type II monitors. Essentially, many type II monitors gather the same signals as type I monitors.

** May have more than three channels, provided that criteria for type III are not met

***May include monitors that measure signals that are in principle able to identify arousals from sleep.

Other definitions and terms

Scoring of sleep studies

The interpretation of the recording from a sleep study may be:

- *Manual*, i.e., performed by a technologist/sleep expert or a sleep physician using predefined criteria. The interpreter may use proper software that facilitates the display of the signals on a screen, but does not score the recordings.
- *Automated*, i.e., the signal is analyzed during its recording or later, and the interpretation of the recording is performed by specialized software. The software identifies the respiratory events using specialized algorithms (and some pieces of software may also crudely evaluate the actual sleep time; e.g., see Pillar 2003³⁰), which typically vary by the specific make of monitor.
- *Combined automated and manual*, i.e., automated analysis with human supervision. This typically pertains to manual corrections on the results of the automated analysis. An additional scenario is when a technologist selects parts of the recordings for software analysis based on signal quality or absence of artifacts.

Setting

Studies with type I monitors (facility-based PSG) are conducted in specialized sleep centers/laboratories. Sleep studies with portable monitors may be:

- *Home-based*, when the study is conducted at home.
- *Hospital-based* when the study is conducted in a hospital ward that is not part of a specialized sleep clinic or a sleep laboratory
- *Sleep clinic-based, facility-based or sleep laboratory-based*, when the study is conducted in a specialized facility.

Attendance

Sleep studies performed in the hospital or in the sleep lab (with any type of monitor) may be:

- *Attended*, if a technologist supervises the recording during sleep time, and has the ability to intervene if needed. This means that the technologist would correct any mistakes leading to bad quality of signals or loss of signals (e.g., probe displacement, or probe failure).
- *Unattended*, if there is no supervision by a technologist during sleep time. These sleep studies may have been home-based or laboratory-based. We classified all sleep studies with portable monitors that were home-based as unattended, including the ones with a continuous feedback to a sleep technologist in the hospital over a modem connection.³¹ The same was true for monitors that have incorporated self-check algorithms and alert (wake) the evaluated subject if a probe is not properly connected or dislodged.³²

Search Strategy

We conducted comprehensive searches in MEDLINE from its inception through the 28th of February 2007 to identify English language publications that described prospective studies comparing portable monitors with facility-based PSG, or describing adverse events or complications of sleep studies. Because different questions were added at different time points, three electronic searches were performed. Searches were

incremental (i.e., the latest search included all the citations of the previous searches). Relevant systematic reviews and meta-analyses, consensus statements and recommendations were also identified. Various search terms were used, including terms that described sleep studies with different monitors, OSAHS, and continuous positive airway pressure (CPAP) treatment for OSAHS. The complete search strategy is reported in detail in **Appendix A**. Reference lists from relevant systematic or non-systematic reviews, the reviewed studies, and publications on practice recommendations were also examined for potentially useful additional citations.

Because rare adverse events or complications would probably not be described at all in the eligible studies (see below for eligibility criteria), the website of the Center for Devices and Radiological Health of the US Food and Drug Administration (FDA) was searched. Specifically, the database of adverse events secondary to medical device use was queried for all available years (1992-2006) using the general term “sleep stud*” (last search December 12, 2006). The database is publicly accessible at <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfMAUDE/search.cfm>.

Inclusion and exclusion criteria

After consultation with our technical expert, AHRQ and CMS, we considered all research studies published in English that met the criteria described in the following sections.

Participants

Disease-free participants, people suspected of OSAHS, or patients with any diagnosis of OSAHS were eligible, provided that they were adults, medically stable and had not been subjected to OSAHS-related surgery. Studies were eligible irrespective of the potential presence of comorbidities, such as chronic obstructive pulmonary disease and congestive heart failure, racial, or gender distributions. For studies comparing the diagnostic ability of sleep monitors, 11 or more participants had to be analyzed after any exclusions or dropouts. For studies assessing adverse events only, 101 or more subjects had to be included. Studies evaluating the ability of facility-based PSG to predict response to CPAP treatment had to have at least 2 weeks duration of CPAP treatment. These cutoffs are arbitrary, and are set to distinguish studies that are small (and perhaps less likely to be informative). For example the sensitivities and specificities obtained from a diagnostic study on 10 people would be extremely uncertain (would have very wide confidence intervals). Similarly, for adverse events, it is likely that less common adverse events and harms will be observed in studies much larger than 100 participants.

Interventions and assessments

- Any sleep studies that were performed with facility-based PSG or portable monitors:
1. For the assessment of the diagnostic ability of portable monitors, eligible studies comparing measurements with portable monitors and facility-based PSG in the same patients, either simultaneously, or within three months of the two measurements (provided that no active treatment was offered between the measurements). For the assessment of type II monitors versus facility-based PSG, we also accepted studies that compared unattended type II monitors at home with attended type II monitors in the hospital. This exception was allowed because of the relative paucity

of data on type II monitors, and because type II monitors essentially record all the signals that non-portable, facility-based PSG records.

2. Only one monitoring device (either portable monitor or facility-based PSG) was sufficient for:
 - 2.1. The assessment of the relationship between baseline severity of OSAHS (as conveyed by AHI, RDI or other indices obtained from facility-based PSG or portable monitors) and the response to CPAP treatment.
 - 2.2. The comparison of the diagnostic ability of manual versus automated scoring of recordings and errors, adverse events, or complications related to different scoring methods
 - 2.3. Errors, adverse events, or complications related to unattended sleep studies, or sleep studies with different monitors.
 - 2.4. Rates of data loss from sleep studies with different monitors.

All types of devices were eligible. However, especially for type IV devices, we excluded the few studies that did not measure directly at least one respiratory signal or the O₂ saturation. Thus, studies using only static charge-sensitive mattresses, only Holter recordings for heart rate, or studies that used only analysis of snoring sounds were excluded. Similarly, we excluded studies that used pulse oximetry but analyzed only the variability of the heart rate (i.e., used oximetry in lieu of ECG to detect pulse rate) and did not evaluate O₂ saturation patterns. In general, monitors that did not record a respiratory signal or SaO₂ during sleep rely on “indirect” assessment of respiratory disturbances in people suspected for OSAHS, and most often were described in older studies. The frequency of respiratory disturbances is a key issue in the diagnosis of OSAHS, and is assessed by the vast majority of modern monitors.

Eligible outcomes - eligible methods of analyses

Ability of AHI or RDI to predict response to CPAP

Any outcome or any measure of association was eligible.

Assessment of concordance between measurements

The reader is referred to Section A2 for a discussion of appropriate methods to assess the concordance of paired measurements. We a priori decided that eligible methods for the assessment of concordance of the measurements were:

1. Difference versus average analyses using either untransformed measurements or using the logarithmic transformation (the only transformation that has a natural meaning in the context of measurement method comparison). This type of analysis is otherwise referred to as Bland-Altman plots, after the researchers that advocated their widespread use.^{33,34}
2. Lin’s concordance correlation coefficient^{35,36} or intraclass correlation coefficients.
3. Weighted or unweighted κ for agreement.

Moreover, studies reporting sensitivity and specificity pairs, or receiver operating characteristic curve analyses (assuming that measurements with facility-based PSG have negligible error for any practical purposes) were also eligible.

Evaluation of errors, adverse events or complications

Any mention of the above was eligible.

Outcomes or metrics that were not summarized

We did not summarize evidence on the repeatability of measurements with the same device, or assessments of concordance using the Pearson correlation coefficient or ordinary least-square regressions. See Section A2 for the rationale that supports this decision. We also did not review evaluations that utilized respiratory events rather than participants as a unit of analysis.

Design

In principle, comparative trials (randomized and non-randomized) that assessed hard clinical outcomes (mortality, cardiovascular morbidity and similar outcomes) across patients managed using laboratory-based PSG measurements and patients managed using portable monitoring would be eligible, but none was found.

Only prospective studies were assessed. For the questions on the concordance of the measurements with different monitors, we excluded studies with overt verification bias (i.e., studies where only those with high RDI in the portable monitor were assessed with facility-based PSG). We note that this does not mean that all included studies are immune to verification bias, as it is very possible that such selections might have been made but not reported in the published paper.

Data abstraction

All studies were extracted in preconstructed forms. The forms were piloted in a set of 6 studies (two for each type of portable monitor) and corrections were made. Because more questions were added to the technology assessment during the data extraction, the forms were amended, and the already reviewed or excluded papers were re-evaluated for potential eligibility for the added questions. A version of the data extraction form was used for studies that were being evaluated for description of adverse events, complications, or harms. **Appendix B** depicts the final versions of the data abstraction forms. Briefly, we recorded information on the citation, patient sampling, pretest probability of OSAHS, characteristics and details of the sleep monitors used and the setting of their use, agreement between measurements or other results as applicable, information on signal loss, and methodological quality items.

Whenever data were reported in good quality graphs, they were electronically digitized using specialized software (Engauge Digitizer, ver 2.14, Mark Mitchell). The digitized data points were then re-analyzed by the Tufts-NEMC EPC to verify and complement information on the studies' results. This was usually necessary for studies that did not assess the concordance of the measurements using difference versus average plots, but only provided a scatter plot of the measurements with the two compared methods. Because digitizing and re-analyzing is a very time-consuming exercise, 10 figures were not digitized. However, the non-digitized papers already provided in tables and in the text almost all pieces of information needed for the purposes of the technology assessment.

All papers were extracted in duplicate. Duplicate extraction was completely blinded for approximately a third of the papers. A critical corroboration of the abstraction form against the published paper was undertaken in the remaining two thirds of the papers. All included and excluded papers were reviewed in detail by the first author, who was also the arbitrator for potential discordant items.

Analyses

Assessing whether baseline AHI or RDI predicts response to CPAP

The association between baseline AHI or RDI and the response to CPAP or adherence to CPAP use was assessed quite differently in the identified studies, precluding any quantitative synthesis. Any measure of response to CPAP was acceptable, and these included quality of life outcomes, objective, or subjective symptom scores, changes in blood pressure, weight loss or other indices. We accepted only objective methods of quantifying adherence, because self-reporting has been shown to be very inaccurate. Thus, sufficient CPAP use (as defined in each study) had to be documented by reading data from built in covert or overt modules that measured the length of actual use. CPAP use had to be least 2 weeks long in all eligible studies. We did not exclude studies that also used other interventions (e.g., weight loss interventions).

Given the nature of the available data and the confounding by other interventions, it is clear that any findings from these assessments should be considered hypotheses forming and should therefore be interpreted with caution.

Comparison of measurements with different methods

For the comparison of the measurements with portable monitors versus facility-based PSG, we grouped studies by type of portable monitor (II, III, IV), setting (home, hospital ward-non-specialized units, specialized sleep units), and scoring method (manual, combined manual and automated, or automated). Especially for type IV monitors, we also distinguish between type IV monitor that recorded only one or two bioparameters from monitors that recorded more than two bioparameters (see Section A3 for more discussion).

Findings of studies falling in the same groups were evaluated comparatively as described below.

Assessing the agreement between measurements

The concordance of two measurements is commonly assessed by difference versus average plots (Bland-Altman plots— see Section A2 for a description).^{33,34} For each study with available information (either reported in the paper or after figure digitizing), we visually depict the average difference between the two measurements and the spread of the 95% limits of agreement (i.e., the boundaries that include the 95% of the differences between the two measurements) (see Section A2). Note that these visual descriptions do not inform on potential changes in the difference between the two measurements at different levels of the measured quantity. Such dependencies were assessed with the Bradley and Blackwood F test³⁷ for digitized data, or as described in the papers for non-digitized graphs.

Analyses with difference versus average plots assume that none of the two methods is better than the other; they merely assess their concordance. For type IV monitors versus facility-based PSG this assumption may not be strictly true. Theoretically, Type IV monitors utilize only a fraction of the information available in facility-based PSG. This is true for type III monitors also, but for type IV monitors this is even more pronounced.

Assessing the ability of portable monitors to predict AHI values suggestive of OSAHS with facility-based PSG

For these analyses, we assumed that facility-based PSG AHI estimates had negligible error for any practical purposes. A detailed description of quantities that were used in the assessment of the diagnostic ability of portable monitors (namely sensitivity, specificity and likelihood ratios) is provided in section A2, along with additional methodological and statistical considerations.

Briefly, the sensitivity and specificity of the portable monitors were derived and visually depicted in (square sensitivity/1-specificity) plots. Studies that yielded high positive likelihood ratio and/or low negative likelihood ratio were identified. For operational cutoffs for a high positive likelihood ratio and a low negative likelihood ratio we used the values 10 and 0.1, respectively (see Section A2 for a more detailed description).³⁸

Finally, AHI cutoffs suggestive of OSAHS were defined differently across studies, or even in the same study. Main emphasis was given to the cutoff of 15 events/hour of sleep. This AHI cutoff is part of the Medicare reimbursement criteria that describe eligibility for CPAP treatment, and is usually considered as suggestive of OSAHS by many experts.³⁹ (Also, see results under Aim B, the majority of the studies used 15 events/hour as a cutoff). However, for completeness, findings from studies using AHI cutoffs of 10 and 20 events/hour of sleep were also presented.

Note that summary estimates of sensitivity and specificity from separate meta-analyses and summary ROC analyses were not presented. It is arguable whether it is meaningful to provide any summary estimates even for monitors that belong to the same type (and have even been studies in the same setting) because quite different definitions for RDI have been used. The qualitative description of the studies' findings was deemed sufficiently informative.

Comparison of rates of data loss, errors, harms, adverse events and complications

Errors resulting in data loss during sleep studies were differently defined across studies. For operational purposes, we adopted each study's definition of an unsatisfactory sleep recording. Because of the variability and the inconsistency of reporting, only qualitative assessments were performed.

There is no clear definition of harms, complications, or adverse events. Thus, any information that was deemed relevant was recorded. Differences in definitions across studies are unavoidable. Therefore, only qualitative descriptions of these rates were performed.

Assessing individual studies

Applicability and quality of studies

The applicability of individual studies to the Medicare population of interest was assessed based on baseline severity of OSAHS (as conveyed by the mean AHI with facility-based PSG), the gender distribution of the participants, their mean age and body mass index, the potential presence of comorbidities, and the selection criteria of the individual studies (sampling population).

We did not assess the quality of studies that reported errors, adverse events, harms or complications. Any information on harms is important, irrespectively of study design and overall methodological quality. According to the inclusion criteria, all studies were prospective and without overt verification bias. Thus, we further assessed the

methodological quality of the studies comparing measurements with different methods based on several quality items: blinding of assessors to results of the other test, blinding to clinical information, enrollment of consecutive patients, random order of measurements or simultaneous measurements with the compared methods, proportion of data loss and clear description of the evaluated population. Quality was evaluated using a three-point scale (A, B, C or good, moderate, poor).

Grade A (good methodological quality) studies fulfill most commonly held concepts of high quality, including the following: blinding of assessors to results of the other test, blinding to clinical information, enrollment of consecutive patients, random order of measurements or simultaneous measurements with the compared methods, clear description of the evaluated population, setting, and measurement methods; appropriate measurement of outcomes; appropriate statistical and analytic methods and reporting; no reporting errors; not excessive data loss (<20%); and no obvious bias.

Grade B (moderate methodological quality) studies may be susceptible to some bias, but not sufficient to invalidate the results. Such studies do not meet the criteria described in category A. They have some deficiencies but none likely to cause major bias. Study may be missing information making assessment of the limitations and potential problems difficult.

Grade C (poor methodological quality) studies are subject to significant bias that may invalidate the results. Such studies may have serious errors in design, analysis or reporting. These studies may have large amounts of missing information or discrepancies in reporting.

Results

Specific Aim A: Clarifications of important topics

The following sections discuss three important issues that impact on the analysis and interpretation of studies comparing portable monitors with laboratory-based PSG.

A1. Is laboratory-based polysomnography considered a “gold standard” for the diagnosis of sleep apnea?

Most experts consider laboratory-based PSG as the reference method for the measurement of AHI during the evaluation of OSAHS.^{1,13} This does not mean that facility-based PSG is an error-free “gold standard” for the diagnosis of OSAHS. An error-free “gold standard” would be a set of criteria or measurements that distinguish subjects with OSAHS from those without, with small misclassification errors for any practical purpose. The error-free “gold standard” thus would have inherent prognostic ability, because subjects with OSAHS have different prognosis from subjects without OSAHS.

There are two complementary lines of reasoning that can be used to challenge the role of facility-based PSG as a necessary and sufficient “gold standard” for the diagnosis of OSAHS. The first pertains to whether sleep studies can actually measure AHI reliably; the second pertains to whether information derived from sleep studies correlates with symptom severity or prognosis.

Does laboratory-based PSG measure AHI reliably?

Here we should distinguish between the technical properties of the probes that are used to record the various cardiorespiratory and neurophysiologic signals, the interpretation of the recordings, and the factors that may affect sleep quality during the sleep study.

Technical properties of the probes

Airflow measurements: Airflow disturbances are usually assessed with thermistors, nasal cannulae (nasal pressure transducers) or end tidal CO₂ monitors in laboratory-based polysomnography. In the recently released AASM Scoring Manual⁴⁰, it is recommended that both an oronasal thermistor and a nasal cannula pressure transducer be used to maximally detect both apneas and hypopneas. Thermistors tend to underdiagnose hypopneas (therefore the nasal pressure transducer is needed) and the nasal pressure transducers tend to overdiagnose apneas (therefore the thermistors are needed).

Thermistors are resistors with the ability to change their electrical resistance with temperature. Thermistors have a monotonic response to temperature changes, which is known for each kind of thermistor in various temperature ranges. However, temperature changes and airflow changes are not linearly linked. Therefore, thermistors may not be able to reliably identify hypopneas. In fact, the American Academy of Sleep Medicine recommended against using only thermistor probes for polysomnography purposes.⁴¹ Thermistors can be used to assess oronasal flow (oronasal probes).

Nasal cannulae measure changes in pressure, and can provide a linear approximation of airflow (after a mathematical transformation of the signal⁴¹). However, they do not

directly assess oral flow. Thus, for people who are mouth-breathing for a long period of their sleep, nasal cannula signal would be largely non-informative. Additionally, intermittent mouth breathing could be misdiagnosed as respiratory events.

Oxygen saturation: It is known that the sampling rate of the oximeters affects the frequency of artifacts and the accuracy of their signal. Data storage every 12 seconds has been claimed to be adequate to identify desaturations with few artifacts in a study.⁴² However, in the same study differences up to 8.4 desaturations/hour on average were identified when data were being stored every 2 seconds compared to every 12 seconds.⁴² Different oximeters may have very different sampling rates (up to as fast as 10 times per second). Moreover, the oximeters that are used in the PSG studies may report their signals differently. Some oximeters report an average of all samplings over a brief sampling period (2 to 12 seconds), whereas others provide “continuous” measurements. An example that highlights the role of oximetry in polysomnography is the Zafar 2005 study.⁴³ This study assessed 113 people with laboratory-based PSG using four different oximeters to evaluate desaturations (hypopneas were defined as reductions in airflow associated with O₂ desaturations). In three oximeters, 35 people would be classified in the group with AHI≤15 events/hour; in the fourth oximeter, 7 of those 35 would be classified in the group with AHI>15 events/hour.

Other probes: Similar concerns probably apply to other probes as well (thoracoabdominal bands, respiratory inductance plethysmography, electroencephalogram, etc.). For example, the best method to detect effort is an esophageal manometer or an inductance plethysmograph. Other methods are not considered to be as accurate.⁴⁰

As evident from the discussion above, different sensors are expected to give different measurements if they were to be applied simultaneously to the same patient.⁴⁴

Interpretation of laboratory-based PSG

Intra- and inter-rater agreement: It is likely that a large proportion of patients assessed for OSAHS have AHI in the vicinity of 15 events/hour, the usually employed cutoff to characterize the presence of the condition. In fact in 29 of the 95 studies that were eligible for Aim B we were able to calculate the percentage of participants who had AHI in lab-based PSG between 10 and 20 events/hour: the median was 18% (interquartile range: 10, 21% of participants). Thus, for the same sleep study even small variations in repeated scoring with the same interpreter (intrarater variability) or with different scorers (interrater variability) might not be completely negligible.

In a study of 11 technologists from nine sleep laboratories significant inter-rater variability was present in scoring of both sleep and respiratory events, and more variability was observed for respiratory events scoring.⁴⁵ In various studies the epoch-by-epoch agreement in the identification of respiratory events is relatively high (kappa for concordance 0.82,⁴⁶ and comparable in other publications^{47,48}). In the Sleep Heart Health Study, standardized interpretation of recordings and standardized scoring criteria resulted in high inter-scorer agreement.⁴⁸⁻⁵⁰ Overall, agreement seems to be better for observers who belong to the same team (same clinic) and for subjects with minimal sleep fragmentation.⁴⁴

Concordance between different scoring methods: The concordance of the measurements also depends on the scoring method. Manual scoring and combined

manual and automated scoring of the same facility-based PSG measurements may yield different results.⁵¹⁻⁵⁷ Similar considerations pertain to differences in the scoring criteria. For example different definitions of hypopnea in PSG (e.g., airflow reduction associated with arousals or desaturations versus airflow reduction associated with desaturations only) are expected to result in different AHI measurements.⁵⁸ Pittman 2004⁵⁹ scored respiratory events according to the American Academy of Sleep Medicine criteria (“Chicago criteria”)⁴¹ and according to the “Medicare criteria” (term used in the study) for 29 people.⁶⁰ For that study, we calculated a kappa of 0.28, indicative of suboptimal concordance using a cutoff of 15 events/hour for the diagnosis of OSAHS. Moreover, we calculated on average 13 events/hour more when the “Chicago criteria” were used (95% limits of agreement: -1, 27; based on our own analyses of the individual participant data, as provided in the pertinent publication⁵⁹).

Sleep quality during the sleep study

It has been advocated that the quality of sleep during the sleep study is adversely affected, because the evaluated subject is hooked up with multiple probes and knows that he/she is being evaluated. The existence of a “first night effect” has been hypothesized because of the lack of familiarity with the sleep study procedures.^{61,62} It is postulated that sleep quality will improve once people become familiar with the sleep-study procedures. However, a first night effect was not documented in repeated home-based measurements in the Sleep Heart Health Study.⁶³ Finally, it has been hypothesized that sleep quality in an unfamiliar environment like the sleep clinic or the sleep laboratory is not optimal and may be improved by pharmacological intervention.⁶⁴

Night-to-night variability and variability across different laboratories

The repeatability and reproducibility of PSG measurements should also be considered. Repeatability (the agreement of serial PSG measurements in the same patient in the same laboratory) and reproducibility (the agreement of PSG measurements in the same patient across different laboratories) of PSG may result in differential classifications of evaluated subjects, if their AHI values are in the vicinity of the cutoff threshold. This is probably true for many people evaluated in the sleep centers and sleep laboratories. Usually the reproducibility of measurements is expected to be worse than their repeatability, because more sources of variation are introduced when measurements are done in different laboratories. In the population setting, the reproducibility of the measurements is of most important, given that subjects will be evaluated in different laboratories.

Is AHI (or RDI) sufficient to diagnose OSAHS?

Facility-based PSG and portable monitoring do not inform on aspects of OSAHS other than the measured sleep parameters. It is acknowledged that the AHI does not correlate well with the intensity of the symptoms in patients with OSAHS.¹³ The correlations between AHI (and other PSG indices such as arousals or desaturation variables) and daytime measures of quality of life, well-being, subjective sleepiness, symptoms and cognitive performance are weak.⁶⁵ There are probably no clinical or statistical differences between patients who differ only by a few points in the AHI.

Moreover, as discussed in Section B1,⁶⁶⁻⁷³ AHI is not well correlated with response to CPAP therapy, or compliance to the therapy itself, among people selected for CPAP

treatment. These findings may imply that the increased accuracy in the measurement of AHI obtained by facility-based PSG may not be predictive of adherence to therapy and therefore prognosis (see section B1 for details).⁷³

Thus, polysomnographic indices alone are not sufficient to classify people into those with and without OSAHS. This is reflected in the design of the studies included in this review, where many different thresholds of AHI have been used as suggestive of OSAHS. Actual thresholds range broadly from AHI \geq 5 events/hour to \geq 40 events/hour. Similarly, this is also acknowledged in the Medicare reimbursement criteria that describe eligibility for CPAP treatment, where a composite definition is employed: AHI \geq 15 events/hour irrespectively of symptoms, or AHI \geq 5 events/hour with symptoms.

Is AHI necessary for the management of people suspected for OSAHS?

As of this writing, no RCT were identified that compared hard clinical outcomes between people managed with facility-based PSG and with portable monitors only. However, a recent RCT by Mulgrew 2007⁷⁴ evaluated the utility of a diagnostic algorithm that did not involve facility-based PSG in the initial management of people suspected for OSAHS. In brief, 68 patients with high probability for AHI >15 events/hour (i.e., moderate to severe OSAHS) were selected on the basis of Epworth Sleepiness Scale score, Sleep Apnea Clinical Score and overnight oximetry that were suggestive of OSAHS. They were randomly assigned to CPAP titration guided by facility-based PSG or ambulatory CPAP auto-titration (without facility-based PSG). The latter arm used a combination of auto-CPAP and overnight oximetry. The population enrolled in the Mulgrew 2007⁷⁴ RCT is only a very small fraction of the total number of people screened for participation.

After 3 months there were no differences between arms in the AHI on CPAP (the primary endpoint): the average difference in AHI on CPAP was 0.8 events/hour (95% CI: -0.9, 2.3). Both arms achieved low median AHI on CPAP at three months (median 3.2 versus 2.5 events/hour in the arms that used and did not use facility-based PSG, respectively). No differences beyond chance were found for the secondary outcomes of the RCT. The difference in the change from baseline in the Epworth Sleepiness Scale score was 1 (p=0.26 for the between-arm comparison). The corresponding difference for the Sleep Apnea Quality of Life Index was 0.17 (p=0.69 for the between-arm comparison). Scores for both aforementioned secondary outcomes improved in all patients compared to baseline, with the exception of a single participant with Cheyne-Stokes respiration. Finally, adherence to CPAP was higher (p=0.021) in the arm that did not receive facility-based PSG (median CPAP use, 6.0 hours/night [interquartile range: 5.1, 7.1]) compared to the arm that received facility-based PSG (median use, 5.4 hours/night [interquartile range: 3.7, 6.4]).

The RCT concluded that in the initial management of patients with *high probability of OSAHS*, PSG testing confers no advantage over an ambulatory approach in terms of diagnosis and CPAP titration. There was also evidence that adherence was better with the ambulatory approach. These findings are in accordance the findings of section B1 (baseline values in metrics obtained from facility-based PSG do not correlate well with response to CPAP or adherence to CPAP use). The Mulgrew 2007 RCT⁷⁴ was not eligible for section B1 because it did not fulfill the pertinent inclusion criteria (correlations/associations of baseline measurements with outcomes were not reported).

How should portable monitors be evaluated?

As discussed in the previous paragraphs, polysomnographic indices alone are not a “gold standard” for the *diagnosis* of OSAHS. However, clinicians consider that PSG information is important in the management of patients with disturbed sleep.¹³ So how should portable monitors be evaluated?

From a pragmatic point of view, one should evaluate clinically meaningful health outcomes in people who have been managed based on information obtained with different sleep monitors. As of this writing no studies have contrasted patient management based on portable monitors versus facility-based PSG with respect to hard clinical endpoints (such as mortality, cardiovascular disease, etc.).

Given the paucity of studies comparing different management options, it is logical to assess the concordance of AHI and RDI values obtained from portable monitors and laboratory-based PSG. Moreover, one may assess whether portable monitors agree with facility-based PSG in classifying subjects above or below a given AHI threshold. This was the design of almost all studies included in the quantitative part of this technology assessment (Section B).

A2. Appropriate methods for the comparison of diagnostic test performance for obstructive sleep apnea

This section provides a discussion of the design and analysis of studies comparing measurements obtained from sleep recordings.

Study design issues

Comparing management based on information from different types of sleep monitors

From a pragmatic point of view, one would like to evaluate hard outcomes in subjects who were managed based on information from different sleep monitors. Typically, hard outcomes include mortality, cardiovascular morbidity, incidence of accidents, and similar endpoints. A randomized controlled trial (RCT) would be the ideal study design.

However, there are great practical and logistical difficulties. For hard clinical outcomes, a study would probably have to evaluate a large number of participant-years to be adequately powered to detect differences (or to exclude clinically meaningful differences).

Alternatively, surrogate endpoints and soft outcomes may be chosen, but their interpretation may be challenging.⁷⁵ Surrogate endpoints are endpoints that may correlate with a hard clinical endpoint. Not all surrogate outcomes are valid,⁷⁵ since interventions that inflict favorable changes in them may not result in corresponding favorable changes in hard outcomes. Soft outcomes, such as self-reported quality of life or symptom scores, may be susceptible to biases, and therefore may not be reliable and easy to interpret.

We identified no randomized trials or non-randomized comparative studies assessing hard clinical outcomes. We identified a single randomized trial that allowed (among others) the evaluation of changes in the Sleep-Apnea Quality of Life Instrument, a subjective and soft endpoint (see Section B1).⁷³

Assessing concordance of measurements obtained with different types of sleep monitors

Most studies assess the concordance of AHI or RDI measurements obtained with different monitor types. This design does not allow inferences with respect to clinical outcomes.

Here, it would be optimal to know the variability in the measurements with the two methods. Therefore, ideally, at least two measurements with each method should be taken. This is even more important when the two sleep monitors were used on different nights.

Issues on statistical analyses in the comparison of measurements with two methods

In the particular case of AHI measurements with portable monitors and with facility-based PSG, one can use two conceptually different analytic strategies:

1. How much do the *individual measurements* differ? Here one quantifies the extent of the *agreement* between the individual measurements. As discussed in the following paragraphs, this is the type of comparisons that are often mistakenly and misleadingly analyzed in the various studies.
2. Does the portable monitor agree with the reference standard (facility-based PSG) in *classifying people into categories that are clinically meaningful* (i.e., suggestive of OSAHS or not)?

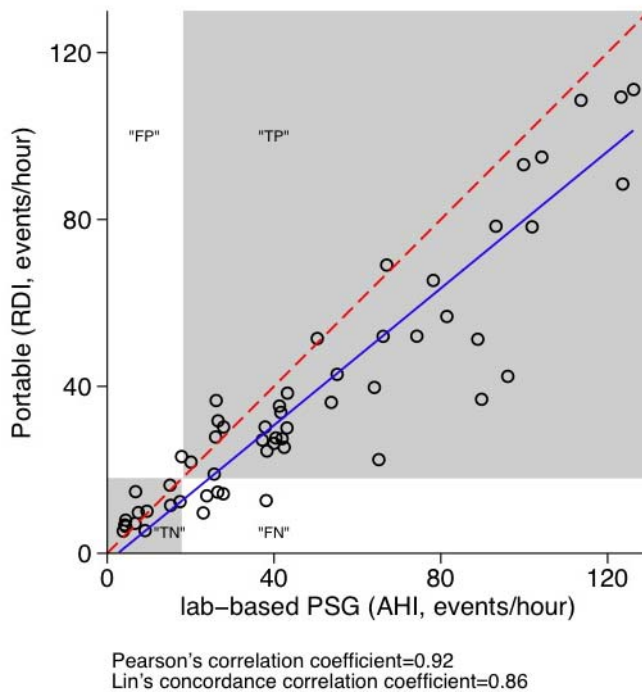
These two questions are related but conceptually different. The first question assesses whether the measurement methods are interchangeable or not. This question would be asked if the actual values of the AHI measurements were of interest. The second question assesses the ability of the different monitors to provide a binary “diagnosis”. Note that two monitors may agree in a binary classification, although their actual measurements may differ substantially. This is illustrated in the worked example that follows.

The following section is a re-analysis of an eligible study that assessed type IV monitors; the individual participant data from this study were digitized by the EPC. We present it as a case study and proceed to discuss some statistical issues in this specific example.

First approach: Assessing the agreement of individual measurements

Figure 2 illustrates measurements in the AHI from the study by Ayappa 2004⁷⁶ (digitized data, re-analyzed by the EPC). The measurements were not performed simultaneously, and the night-to-night variability of the sleep monitors was not described. Ayappa 2004 considered an $AHI \geq 18$ events/hour as suggestive of OSAHS (other thresholds were also assessed).

Figure 2. Illustrative example: measurements with facility-based PSG and type IV monitors in Ayappa 2004 (digitized data).



FN: false negative; FP: false positive; PSG: polysomnography; TN: true negative; TP: true positive. Digitized data from Ayappa 2004, where type IV monitors (Pro-Tech PTAF2 and Compumedics P2) were compared with facility-based PSG.⁷⁶ Measurements were not taken simultaneously. The dotted line is the line of identity. The solid line is the line that best describes the agreement between the two measurements (derived from a reduced major axis regression; this is different from an ordinary least squares regression, see text for details). The graph is divided in four regions with different shading, according to whether they agree or disagree in classifying people into four categories using 18 events/hour as a cutoff for both measurements (following what was done in the original publication). Assuming that AHI>18 events/hour are diagnostic of OSAHS, facility-based PSG has negligible measurement error, and that night-to-night variability is negligible for practical purposes, the portable monitor would yield “true positive” measurements in 38 cases, “true negative” in 12 measurements, “false positive” in five cases, and “false negative” in one case.

If the two monitors estimated exactly the same frequency of respiratory events, all points in the graph would align on the dashed diagonal line, the *line of identity*. However, the portable monitor tends to yield lower values compared to facility-based PSG. The line that describes the relationship between the two measurements is derived from a proper regression, and is called the *reduced major axis*.⁷⁷ This is different from an ordinary least squares regression, in that it allows both measurements to be *imperfect*. Ordinary least squares regression ignores the error in the horizontal axis (assumes that the measurement in the horizontal axis is *perfect*), which may not be true for many measurement comparison studies.⁷⁷ Therefore, the reduced major axis describes the same relationship between the measurements, even if the horizontal and vertical axes are swapped. On the contrary, ordinary least squares regression may lead to different conclusions when axes are swapped.

Other types of regressions that allow both measurements to be imperfect exist. Examples are Deming regression,⁷⁸ Passing-Bablok regression,⁷⁹ and other methods, whose description is beyond the scope of this section.

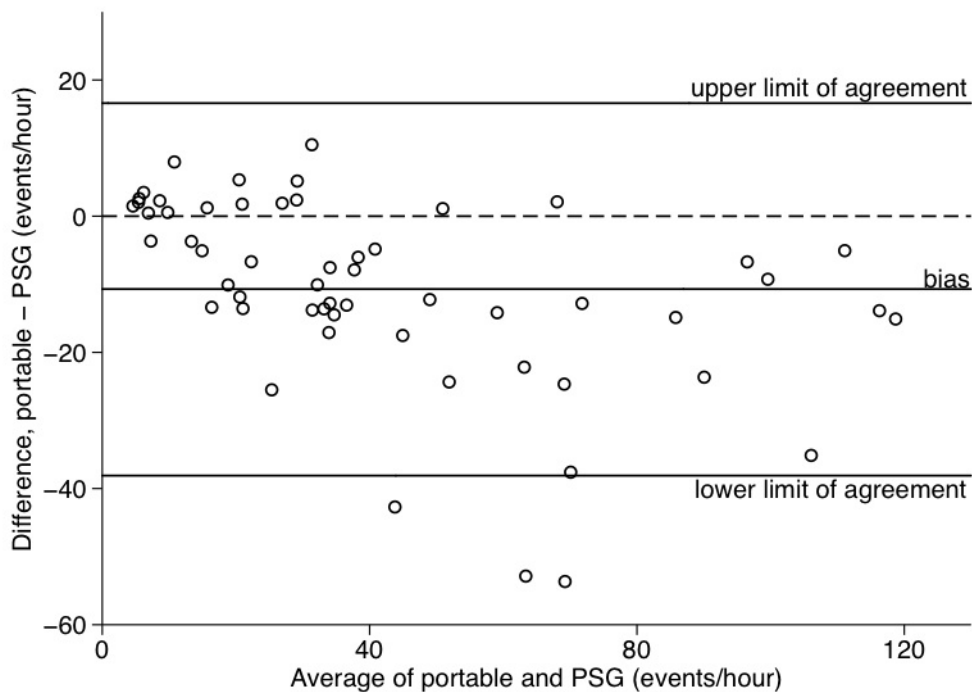
Why correlation analyses are not sufficient

Pearson's correlation coefficient (ρ) describes whether the measurement points lie along any straight line in the plot. In other words Pearson's ρ informs only on the *precision* of the measurements (i.e., how tight the points are scatter along their line of agreement). It does not inform on the *accuracy* of the measurements (i.e., how close the measurements' line of agreement is to the diagonal line of identity). In the Ayappa 2004 example (**Figure 2**) Pearson's ρ was 0.92. An improved metric, Lin's concordance correlation coefficient (ρ_c), penalizes Pearson's ρ according to how far the reduced major axis and the line of identity are.^{35,36} In this example, ρ_c was 0.86. Note that, in measurement comparison studies, high values of ρ or ρ_c do not necessarily indicate high degree of agreement. In general, correlation analyses are not very informative.

Recommended analyses: Difference versus average plots

It is useful to assess the difference in the two measurements for different levels of AHI. Plotting the difference in the two measurements (a measure of their discrepancy) against their average (the best estimate of the unobserved true value) is often very informative.^{33,34} This type of analysis is otherwise referred to as Bland-Altman plots, after the researchers that advocated their widespread use. **Figure 3** shows the Ayappa 2004 example.

Figure 3. Illustrative example: difference versus average analyses of measurements with facility-based PSG and type IV monitors in Ayappa 2004 (digitized data).



Digitized data from Ayappa 2004, where the type IV monitors (Pro-Tech PTAF2 and Compumedics P2) were compared with facility-based PSG.⁷⁶ The dashed line at zero difference is the line of perfect agreement. The mean bias stands for the average systematic difference between the two measurements. The 95% limits of agreement stand are the boundaries within which 95% of the differences lie. If these are very wide and encompass clinically important differences, one may concur that the agreement between the measurements is suboptimal.

Note that the spread of the differences increases for higher values of AHI or RDI. This indicates that the mean bias and 95% limits of agreement do not describe adequately the differences between the two measurements; differences are smaller for smaller AHI or RDI levels and larger for larger AHI or RDI levels. In this example bias = -11 events/hour (95% limits of agreement: -38, 17), with statistically significant dependence of difference on average (Bradley-Blackwood F test, $p < 0.01$).

The difference versus average plots can describe the magnitude of the differences at different AHI or RDI values. Provided that the differences are *evenly scattered* across the plot, their mean value is a useful descriptive metric and represents the systematic bias between the two measurements. The 95% limits of agreement (95% limits of agreement; **Figure 3**) define the region in which 95% of the differences are expected to fall. When the 95% limits of agreement are very broad, the agreement is suboptimal.

However, in the Ayappa 2004 example (**Figure 3**) the differences are not evenly scattered across the plot (i.e., *heteroskedasticity* exists), and this is formally statistically significant (Bradley-Blackwood³⁷ F test, $p < 0.01$). This means that the mean bias and 95% limits of agreement may be misleading, because they do not apply to all AHI or RDI levels. In such cases, the difference may be tolerably small for small levels of AHI or RDI, which are of most interest. Large differences in AHI or RDI when both measurements are very large may not bear particular importance for practical purposes. Therefore, the portable monitor may agree with facility-based PSG in distinguishing “large” from “small” AHI values reasonably well, despite their large differences in the actual individual measurements.

One may transform the measurements in order to overcome issues related to heteroskedasticity of the differences (i.e., the uneven scattering along the plot). Only transformations that have a natural meaning should be used.^{33,34} The logarithmic transformation is probably the only option.^{33,34} In this report, we did not employ transformations in the measurements when difference versus average analyses are presented; however we caution on the interpretation of difference versus average analyses when differences are unevenly distributed in the Bland-Altman plot.

Second approach: Assessing the concordance in classifying people into categories

As discussed in Section A1, facility-based PSG may not be a sufficient reference standard for the *diagnosis of OSAHS*. Briefly, this is because of the night-to-night variability in AHI, the intrarater and interrater variability for any given recording, and because AHI alone does not correlate perfectly with symptoms and quality of life outcomes. Instead, one may assess the ability of portable sleep monitors to predict *an AHI index suggestive of OSAHS* in facility-based PSG (i.e., to predict AHI larger than a cutoff in facility-based PSG).

The ability of the portable monitors to predict an AHI that is suggestive of OSAHS in facility-based PSG (the reference test) is quantified by calculating their *sensitivity* and *specificity* (**Table 3**). Sensitivity is defined as the proportion of people with AHI suggestive of OSAHS that are correctly identified as such by the portable monitor. Specificity is the proportion of people with AHI non-suggestive of OSAHS that are correctly identified as such by the portable monitor. Sensitivity and specificity range between 0 and 100% and higher values imply better diagnostic ability.

Table 3. 2 by 2 table used in the calculation of sensitivity and specificity of portable monitors to predict AHI suggestive of OSAHS in facility-based PSG.

		Facility-based PSG	
		Suggests OSAHS*	Does not suggest OSAHS*
Portable monitor	Suggests OSAHS	"TP" (=38)	"FP" (=5)
	Does not suggest OSAHS	"FN" (=1)	"TN" (=12)

* The assumption is that facility-based PSG has negligible misclassification rates.

From this table, Sensitivity=TP/(TP+FN) and Specificity=TN/(TN+FP)

"TP": true positive; "FN": false negative; "FP": false positive; "TN": true negative. In the parentheses are the corresponding numbers for the worked example (Ayappa 2004⁷⁶). The quotation marks are retained to stress the assumption that facility-based PSG is has negligible misclassification rates for practical purposes. If this assumption does not hold, then the estimates of Sensitivity and Specificity from the above formulae are biased (upwards or downwards, depending on other considerations). See also **Figure 2**.

A particularly informative graph plots the sensitivity of portable monitors against 100% minus their specificity in a square plot (commonly known as the ROC space plot). The closer a study point is to the upper left corner of the plot, the better its diagnostic ability (**Figure 4**).

There are additional ways to assess the ability of portable monitors to predict facility-based PSG results. Each test result (here the test is the portable monitor) changes the certainty of a diagnosis (here the "diagnosis" would be the classification of AHI with facility-based PSG into two categories: suggestive or non-suggestive of OSAHS). For example, when the portable monitor finds an AHI suggestive of OSAHS, one's certainty for a positive diagnosis increases. Reciprocally, when a portable monitor finds an AHI not suggestive of OSAHS, a negative diagnosis becomes more certain. The greater the change in the certainty of diagnosis, the more informative the use of the portable monitor.

The *positive* and *negative likelihood ratios* (LR+ and LR-, respectively) quantify the change in the certainty of the “diagnosis” conferred by the results of the portable monitor. More specifically, the likelihood ratios transform the *pretest odds* to the *posttest odds* of a given (positive or negative) diagnosis:

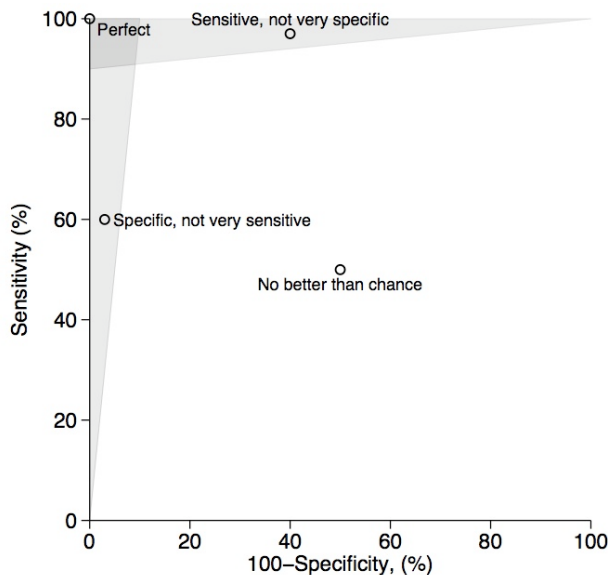
$$posttest\ odds = pretest\ odds \times LR$$

For a positive result with the portable monitor, the positive likelihood ratio would be used in the above relationship; for a negative result with the portable monitor, the negative likelihood ratio would be used. The likelihood ratios can be conveniently calculated as follows:

$$LR+ = \frac{sensitivity}{1 - specificity}, \quad LR- = \frac{1 - sensitivity}{specificity}$$

If a given portable monitor has very good ability to predict the results of facility-based PSG, its positive likelihood ratio will be high (will greatly increase the odds of a positive diagnosis) and its negative likelihood ratio will be low (will diminish substantially the likelihood of the positive diagnosis). A completely non-informative portable monitor would have likelihood ratios equal to 1 (does not transform the pre-test odds substantially in the equation above). Typically, a positive likelihood ratio of 10 or more and a negative likelihood ratio of 0.1 or less are considered to represent informative tests.³⁸ We should note that other, more lenient boundaries for LR+ and LR- can be used^{38,39} and that the choice of the boundaries is a subjective decision. It is interesting to note that studies with high LR+ and low LR- can be readily identified in the square sensitivity/100%-specificity plot, as shown in **Figure 4**.

Figure 4. Square plot of sensitivity versus 100%-specificity.



Four hypothetical studies are depicted in the square sensitivity/100%-specificity plot. The closer a study is to the upper-left corner of the plot, the better its diagnostic ability. Studies lying on the major diagonal of the plot have no diagnostic ability (no better than chance). Studies lying on the left shaded area have positive likelihood ratio (LR+) of 10 or more. Studies lying on the top shaded area have negative likelihood ratio (LR-) of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both LR+>10 and LR-<0.1.

A3. Classification of portable sleeping monitoring devices

As previously listed in **Table 2**, the then “American Sleep Disorders Association” (now American Academy of Sleep Medicine) classified the devices that have been used in sleep studies into four categories, based on the signals they record.²¹ Facility-based PSG is considered a type I device, and is the most comprehensive. The recorded channels allow the identification of respiratory events and sleep staging, and allow the calculation of AHI.

Type II devices record all the signals recorded by Type I devices using a minimum of seven channels (**Table 2**). Type II devices may measure heart rate instead of recording a complete ECG signal according to the classification used. Many type II monitors record practically the same information as type I monitors. Type II monitors allow the measurement of AHI. Type II monitors that record fewer channels than laboratory-based PSG (e.g., seven versus ~14-16) allocate fewer recording channels to certain signals (such as cerebral activity signals).

Type III devices record a minimum of four channels, including two respiration channels, ECG or heart rate, and oxygen saturation. One respiratory channel could be used for airflow monitoring and the other could be used to monitor respiratory movements. By definition, these monitors do not provide information on sleep staging and therefore cannot measure AHI. Instead, they measure RDI, a proxy for AHI.

Type IV devices record even less information. In the ASDA classification type IV monitors record one or two bioparameters.²¹

There is a progressive loss of information on sleep parameters from type I and II monitors to type IV monitors. Information on sleep staging is lost in Type III monitors; and in addition, (some) information on airflow is lost in type IV monitors.

Comment

We should caution that there are some difficulties with the aforementioned classification scheme, which cannot explicitly classify all existing portable monitors. This is especially true for newer portable monitors that measure bioparameters proposed in later years. For example, WatchPAT is a monitor that records heart rate, oximetry, wrist actigraphy and peripheral arterial tonography (a measure of sympathetic activation⁸⁰). WatchPAT would remain “unclassified”: it measures more than two bioparameters (so is not, strictly speaking, a type IV monitor under the original ASDA classification), but does not have at least two airflow channels and cannot be considered a type III monitor under the original classification. As mentioned in the Methods section, for operational purposes we broadened the definition of type IV to include all monitors that *do not meet the criteria for type III devices* (despite the fact that they record more than two bioparameters). This was followed in previous systematic reviews as well.¹³ However we report separately type IV monitors that record only one or two bioparameters from those that record more than two.

A3a. Different parameters of sleep and diagnosis of OSAHS

The assessed studies did not provide much insight on the exact contribution of each of the recorded signals to the diagnosis of OSAHS. We identified only two studies that allowed for a direct assessment of including data gathered from the airflow channels. Using a type IV monitor, Baltzan 2000 found that the addition of the airflow signal

(thermistor) to the oximetry readings provided useful additional information (according to the scorers' judgment) for the interpretation of the record in one-third of 97 participants.⁸¹ Similarly, Gugger 1995⁸² estimated a sensitivity and specificity of 82% and 90%, respectively, to predict AHI >20 events/hour with facility-based PSG using all signals (oximetry and airflow channels) recorded by a type IV monitor (Autoset, diagnostic mode). When only oximetry data were assessed, the corresponding values dropped to 76% and 69%.⁸²

In addition to the above caveats, more general comments can also be made.

Differences between type I/II and type III monitors are obviously expected when the total sleep time is substantially different compared to the total recording time. Even if the same number of respiratory events were detected by both type I/II and type III monitors, type III monitors would yield lower frequency of respiratory events. This is because they overestimate the total sleep time (and hence $RDI \leq AHI$). The discrepancy between the total recording time and the total sleep time increases with the severity of OSAHS (patients with more severe OSAHS have longer cumulative arousals, and thus more Wake After Sleep Onset [WASO]). Moreover, this is true for people who just happened to sleep for a short time during the sleep study. For example, it has been hypothesized that the quality and length of sleep may be adversely affected during a subjects' first sleep study ("first night effect").⁶² However, a first night effect was not documented in repeated home-based measurements in the Sleep Heart Health Study.⁶³

Here we should note that some type IV monitors that record more than three bioparameters, evaluate signals that were proposed after the ASDA classification scheme was developed. For example, peripheral arterial tonometry is a measure of sympathetic activity, and may be a surrogate of airflow disturbances and microarousals.^{30,59} The same caveats that apply to type III monitors, may apply to this category of monitors (type IV that record three or more bioparameters) as well.

The distance between type I/II and type IV that record only one or two bioparameters monitors is even greater. In the latter both the sleep duration and the respiratory events are approximated because they are not measured directly. Given that the majority of evaluated patients would have AHI in the neighborhood of 15 events/hour, it is expected that the misclassification with type IV monitors will be greater compared to type III monitors.¹³

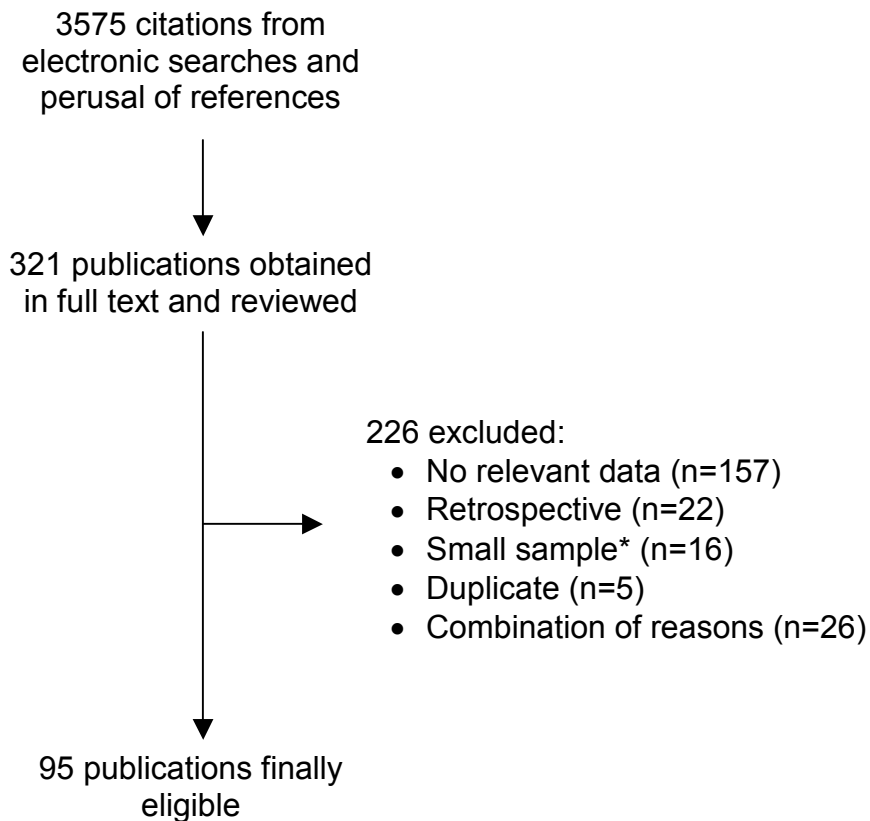
Specific Aim B: Role of portable monitors vs. facility-based polysomnography in the diagnosis of OSAHS

This part of the technology assessment was based on a systematic review of the literature. The presentation is structured so that each section provides the results of the systematic review on the pertinent key question.

Literature flow

Figure 5 shows the number of screened, reviewed, and eligible publications in this report. Overall, 321 papers were reviewed in full text. Finally, 95 studies were included.

Figure 5. Literature flow



* Small sample means 10 or less for comparisons of portable monitors with laboratory-based PSG; or 100 or less for studies with no comparative data with potential information on errors, complications or adverse events of sleep studies (see Methods).

All 95 finally eligible papers pertained to studies providing non-overlapping information.

B1. Ability of facility-based polysomnography vs. portable monitors to predict response to CPAP and changes in clinical outcomes after CPAP treatment

We did not identify any trials that assessed hard clinical outcomes (such as mortality, cardiovascular morbidity, etc.) across patients managed using laboratory-based PSG measurements and patients managed using portable monitoring. We identified eight studies⁶⁶⁻⁷³ that associated baseline severity of OSAHS (as conveyed by AHI, RDI or other indices obtained from facility-based PSG or portable monitors) with response to CPAP or adherence to CPAP use. As mentioned in the methods, CPAP use was at least 2 weeks in all eligible studies.

AHI (or RDI) is considered the most important quantity obtained from sleep studies, as discussed in the previous sections. Therefore we discuss AHI (or RDI) separately from other indices in the following paragraphs.

Associations of indices from sleep studies with adherence to CPAP use

Apnea-hypopnea index

Table 4 lists two RCT^{67,73} and three cohort studies^{66,69,72} that provided some measure of association between adherence to CPAP use and baseline AHI values. With the exception of Bennett 1998⁶⁶ and Whitelaw 2005,⁷³ the mean AHI was above 50 events/hour at baseline (i.e., participants had severe OSAHS).

A crossover RCT⁶⁷ compared fixed versus automatically adopted pressure settings for CPAP. The second RCT⁷³ (Whitelaw 2005) evaluated the ability of experienced physicians to predict a clinically significant improvement in the Sleep Apnea Quality of Life Instrument (SAQLI) after 4 weeks on CPAP. Physicians based their predictions on clinical characteristics and information from a sleep study. For the sleep study, patients were randomized to laboratory-based PSG (n=132 analyzed) or a home-based study with a type IV monitor (“Snoresat”, n=156 analyzed). Adherence to CPAP use or extent of CPAP use in these five studies was measured differently, precluding any quantitative synthesis (**Table 4**, outcome column).

Overall, higher AHI at baseline was associated with longer average CPAP use, and better adherence to CPAP use (**Table 4**). Importantly, in a *post hoc* analysis in the Whitelaw 2005 RCT,⁷³ physicians tried to predict the likelihood that a patient would use CPAP at least 4 hours per day on average during the 3 months of follow-up. In the arm where clinical information and information from laboratory-based PSG was available, the area under the curve (AUC) for physician prediction was 0.77 (95% confidence interval: 0.69, 0.85). In the other arm (type IV monitor) the AUC was 0.77 (95% confidence interval: 0.70, 0.85). Similarly, no differences were observed when the baseline AHI and RDI values alone were used as potential predictors of CPAP compliance (**Table 4**) (AUC=0.79 [95% confidence interval: 0.72, 0.87] and AUC=0.78 [95% confidence interval: 0.70, 0.86], respectively). In these analyses people who did not use CPAP at the end of follow-up (quitters, n=19) were excluded. Quitters had lower AHI or RDI values compared to non-quitters.

In three studies^{67,69,72} participants had a mean AHI>50 events/hour of sleep. Also, in the Whitelaw 2005 RCT,⁷³ people who did not use CPAP at all had lower AHI or RDI values and were excluded from these analyses. Thus, one may hypothesize that

differences in baseline AHI have limited ability to predict CPAP use or compliance among people with severe respiratory disturbances.

Table 4. Associations of baseline apnea-hypopnea index with CPAP use.

Author, Year, Country	Design; criteria for CPAP	N _E (N _A)	Participants	Mean baseline AHI (events/h)	Outcome • effect size (95% CI or p value)
Whitelaw, 2005, Canada	RCT; All on CPAP	307 (288)	Random sampling from consecutive patients needing CPAP	23 ^a	Ability of AHI or RDI alone to predict objectively measured mean CPAP use of ≥4h/d, AUC: • Lab-PSG (n=132): 0.79 (0.72, 0.87) • Snoresat ^b (n=156): 0.78 (0.70, 0.86)
Nosedá, 2000, Belgium	Cohort; CPAP if AHI>20 events/h, consent	124 (106)	All people eligible for CPAP who appeared at the sleep laboratory in 12 mo	62	Correlation between the percentage of days that CPAP was used (during at least 4 mo) and AHI at baseline: • 0.15 (p>0.05)
Lloberes, 2004, Spain	Cohort; CPAP if AHI with symptoms or AHI >30 events/h	133 (88)	Consecutive referrals to a sleep unit	63	Odds ratio for objectively measured mean CPAP use ≥4h/d at 12 mo: ^c • AHI <50 : 1.0 (reference) • AHI 50-63: 28.2 (p=0.016) • AHI 64-78: 13.6 (p=0.036) • AHI >78: 2.4 (p=0.364)
Bennett, 1998, UK	Cohort; All on CPAP	41 (40)	Random sampling from sleep clinic referrals ^d	16	Correlation between mean CPAP use after 4 weeks and AHI at baseline: • 0.34 (0.03, 0.59)
d'Ortho, 2000, France	Crossover RCT; All on CPAP	25 (24)	Consecutive people with OSAHS ^e	57	Difference in mean CPAP use among patients with AHI≥60 vs. AHI<60 events/h: • Fixed CPAP: 1.6h (0.3, 2.9) • Auto CPAP: 1.9h (0.6, 3.2)

AHI: Apnea-hypopnea index; AUC: area under the curve; CI: confidence interval; CPAP: continuous positive airway pressure; d: days; h: hours; mo: months; N_A/N_E: Number analyzed/enrolled; NHP: Nottingham Health Profile; OSAHS: obstructive sleep apnea-hypopnea syndrome; PSG: facility-based polysomnography; RCT: randomized controlled trial; RDI: respiratory disturbance index

^a a size-weighted average across the two arms; RDI was treated as AHI for the portable monitor arm

^b "Snoresat" is a type IV monitor, that was used unattended at home

^c Logistic regression adjusted for Nottingham Health Profile (a quality of life scale), Epworth sleepiness score, age, sex, minimal arterial O₂ saturation; the corresponding odds ratios were non-significant at 3 months of CPAP use (not shown).

^d 30 "randomly recruited from the Oxford Sleep clinic", and the remaining 11 were oversampled on the basis of >4% dip in arterial O₂ saturation

^e Clinical suspicion confirmed by laboratory-based PSG

All AHI measurements were performed with facility-based PSG. Ranges for the AHI at baseline were not reported in these studies. Studies are sorted by decreasing number of analyzed people.

Other indices, apart from the apnea-hypopnea index

Three studies^{66,69,72} reported data on associations of indices other than AHI (or RDI) with CPAP use. No meta-analysis was feasible, because of the differences in the definitions in the three cohorts. Among the various indices that were examined, some were positively or negatively associated beyond chance with a metric of CPAP use. We caution that the clinical significance of these associations is unclear. Furthermore,

because adjustments for multiple comparisons were not made, one cannot exclude the possibility of spurious associations due to chance.

In a cohort of 124 people, Nosedá 2000⁷² assessed the correlation between the percentage of days that the CPAP machine was used and a variety of indices (apart from AHI) derived from facility-based PSG at baseline. The only significant finding was a negative association with the proportion of slow wave sleep (Spearman's correlation coefficient = -0.22, $p < 0.05$). Apnea index, hypopnea index, proportion of REM sleep, mean and minimum O₂ saturation during sleep and an index describing the frequency of changes in the sleep stages were not associated beyond chance with the percentage of days that the CPAP machine was used. In addition, when correlations of the aforementioned indices with the mean effective use per day of effective use were assessed, only the indices for the mean and minimum O₂ saturation during sleep were statistically significant (Spearman's correlation coefficient = -0.25, $p < 0.05$). Note that adjustments for multiple comparisons were not made.

Lloberes 2004⁶⁹ assessed in a multivariable logistic regression the ability of various clinical and laboratory parameters to predict average CPAP use of at least 4 hours per night ("good" adherence) after 3 and 12 months. The model adjusted for a variety of factors (i.e., AHI at baseline, the Nottingham Health Profile, total Epworth Sleepiness Scale, age, sex, and minimum O₂ saturation at baseline). The minimum value of O₂ saturation during baseline PSG was not a significant predictor of good adherence. Compared to a minimum O₂ pressure less than 60 mmHg, people with a minimum O₂ saturation between 60 and 79 mmHg had 1.06 times higher odds for good adherence at 12 months ($p = 0.951$), and those with values of at least 80 mmHg 7.77 times higher odds ($p = 0.072$). The corresponding odds ratios for good adherence at 3 months was 0.38 ($p = 0.25$) and 0.47 ($p = 0.40$).

In a cohort of 41 people (Benett 1998⁶⁶) assessed the correlation between various indices measured at baseline PSG and mean CPAP use after 4 weeks. In detail, the rate of O₂ desaturations and the movement event index were positively correlated with mean CPAP use at 4 weeks beyond chance (Spearman's correlation coefficient 0.34 [$p = 0.03$] and 0.39 [$p = 0.01$], respectively). However, mean use of CPAP at 4 weeks was not significantly correlated with the number of arousals during sleep (either 1.5 second microarousals or any ECG-defined arousals), or an autonomic arousal index (defined as increased heart rate over 4 to 45 second intervals). Adjustments for multiple comparisons were not performed.

Associations of indices from sleep studies with response to CPAP in terms of quality of life outcomes

Apnea-hypopnea index

Overall, two studies^{69,73} assessed the ability of baseline AHI (or RDI⁷³) to predict response to CPAP based on a quality of life instrument (sleep apnea quality of life index, SAQLI⁷³ and Nottingham Health Profile⁶⁹).

The Whitelaw 2005 randomized controlled trial⁷³ of 307 patients (**Table 5**) assessed the ability of experienced sleep physicians to predict a clinically significant improvement (1 unit) in the Sleep Apnea Quality of Life Instrument (SAQLI) after 4 weeks on CPAP. The accuracy of physician prediction was comparably suboptimal across both arms, and there was no evidence that laboratory-based PSG was superior to the type IV monitor

(AUC=0.67 [95% confidence interval: 0.60,0.77] and 0.65 [95% confidence interval: 0.56, 0.75] when laboratory-based PSG and “Snoresat”, respectively, were used in addition to clinical information). In the same trial, baseline AHI or RDI alone did not predict improvement of more than one unit in SAQLI (i.e., the minimum clinically meaningful difference).⁷³ The findings were unchanged when people with very high baseline SAQLI were excluded (because of a “ceiling effect” people with very high SAQLI at baseline simply could not improve more).

Lloberes 2004⁶⁹ found no statistically significant changes between patients or within patients in the Nottingham Health Profile across different AHI categories (**Table 5**) over a time period of 12 months of CPAP use. Their multivariate analyses used proper methods to account for repeated measurements in the Nottingham Health Profile (baseline, 3 months and 12 months) (**Table 5** footnote).

Table 5. Associations of baseline apnea-hypopnea index with response to CPAP in terms of quality of life outcomes.

Author, Year Country	Design; criteria for CPAP	N _E (N _A)	Participants	Mean baseline AHI (events/h)	Outcome • effect size (95% CI or p value)
Whitelaw, 2005 Canada	RCT; All on CPAP	307 (288)	Random sampling from consecutive patients needing CPAP	23 ^a	Ability of AHI or RDI alone to predict change of 1 unit in SAQLI, ^b AUC: • Lab-PSG (n=132): 0.59 (0.49, 0.69) • Snoresat ^c (n=156): 0.66 (0.56, 0.75)
Lloberes, 2004 Spain	Cohort; CPAP if AHI with symptoms or AHI >30 events/h	133 (88)	Consecutive referrals to a sleep unit	63	Change in NHP (multiple time points at baseline, 3 mo, 12 mo analyzed with GEE). ^d Between patients (arms): • AHI <50: reference • AHI 50-63: -4.66 (-17.84, 8.53) • AHI 64-78: -9.66 (-23.68, 4.37) • AHI >78: -1.64 (-14.73, 11.45) Within patients: • AHI <50: reference • AHI 50-63: 1.91 (-3.70, 7.53) • AHI 64-78: 0.79 (-3.20, 4.77) • AHI >78: 1.35 (-1.44, 4.14)

AHI: Apnea-hypopnea index; AUC: area under the curve; CI: confidence interval; CPAP: continuous positive airway pressure; GEE: generalized estimating equation modeling; h: hours; mo: months N_A/N_E: Number analyzed/enrolled; PSG: facility-based polysomnography; RCT: randomized controlled trial; RDI: respiratory disturbance index; SAQLI: Sleep-apnea quality of life index; wk: weeks

^a a size-weighted average across the two arms; RDI was treated as AHI for the portable monitor arm

^b 1 unit is the minimum clinically meaningful difference in the SAQLI instrument

^c “Snoresat” is a type IV monitor, that was used unattended at home

^d GEE adjusted for Epworth sleepiness score, sex, age, CPAP use and minimum arterial O₂ saturation

Other indices, apart from the apnea-hypopnea index

Lloberes 2004⁶⁹ assessed changes in the Nottingham Health Profile between and within arms in a multivariate analyses that accounted for repeated measurements. Their analyses accounted for a variety of factors (i.e., AHI at baseline, total Epworth Sleepiness Scale, age, sex, and minimum O₂ saturation at baseline). There were no statistically

significant differences in the Nottingham Health Profile between patients who had minimum O₂ pressure less than 60 mmHg, between 60 and 79 mmHg and at least 80 mmHg (i.e., no between-patient effects). However, when changes in each person's score over time were assessed, people who were in the latter two categories showed statistically significant improvements compared to those with minimum O₂ pressure less than 60 mmHg (p<0.001 and p=0.02 for the categories of 60-79 mmHg and ≥80 mmHg, respectively).

The clinical significance of this observation is unclear.

Associations of indices from sleep studies with response to CPAP in terms of other outcomes

Apnea-hypopnea index

Table 6 summarizes associations with physiological outcomes that were reported in four prospective cohorts.^{66,68,70,71} Bennett 1998⁶⁶ associated higher AHI at baseline with greater improvement in the Epworth sleepiness score (ESS, a subjective score), and in an objective test that measures maintenance of wakefulness (Oxford sleep resistance test, OSLER). However, the corresponding correlation coefficients were only modest in magnitude (**Table 6**). Most of the other reported associations have no obvious clinical significance and were also statistically not significant (**Table 6**).

Other indices, apart from the apnea-hyponea index

Finally, Bennett 1998⁶⁶ found significant correlations between changes in the ESS and a variety of indices that were assessed: microarousals, any ECG arousal, the rate of O₂ desaturations, the movement event index and an autonomic arousal index (defined as increased heart rate over 4 to 45 second intervals) (p-values <0.02 for all these indices). The same was true when they evaluated correlations of the aforementioned indices with changes in OSLER test scores (p<0.006 for all indices). Overall, increased severity of OSAHS (as conveyed by the respective indices) was associated with improvements in the ESS and the OSLER test (negative correlation coefficients for ESS and positive correlation coefficients for OSLER).

We should note that adjustments for multiple comparisons were not performed.

Table 6. Associations of baseline apnea-hypopnea index with outcomes other than those assessed in Tables 4 and 5.

Author, Year Country	Design; criteria for CPAP	N _E (N _A)	Participants	Mean baseline AHI	Outcome and effect size (95% CI or p value)
Hermida, 2004 Spain ^a	Cohort; CPAP if AHI ≥30 events/h or AHI >10 and <30 events/h with symptoms	83 (83)	Referrals to a sleep clinic	57	Correlation between baseline AHI and <ul style="list-style-type: none"> • Change in 24h-mean SBP from baseline: 0.01 (p=0.93) • Change in 24h-mean DBP from baseline: 0.07 (p=0.54)
Nosedá, 1996 Belgium	Cohort; All on CPAP	95 (39)	People needing CPAP referred to a sleep lab ^b	67	Correlation between change in AHI from baseline with AHI at baseline: ^c <ul style="list-style-type: none"> • -0.31 (p<0.05)
Bennett, 1998 UK	Cohort; All on CPAP	41 (40)	Random sampling from sleep clinic referrals ^d	16	Correlation between AHI at baseline and <ul style="list-style-type: none"> • Change in OSLER from baseline to 4 wk: 0.52 (0.26, 0.72)^e • Change in ESS from baseline to 4 wk: -0.53 (-0.72, -0.25)
Marrone 2003 Italy	Cohort; All on CPAP	13 (13)	Consecutive people undergoing PSG for suspected OSAHS	80	Correlation between baseline AHI and <ul style="list-style-type: none"> • Change in the range of SBP during apneic events: ND (p>0.05) • Change in the range of SBP during apneic events: ND (p>0.05)

AHI: Apnea-hypopnea index; CI: confidence interval; CPAP: continuous positive airway pressure; h: hours; DBP: diastolic blood pressure; ESS: Epworth Sleepiness Scale; N_A/N_E: Number analyzed/enrolled; ND: not described; OSAHS: obstructive sleep apnea-hypopnea syndrome; OSLER: Oxford sleep resistance test; PSG: facility-based polysomnography; SBP: systolic blood pressure; wk: weeks

^a Patients are probably the same as those described in Zamarron 1999 and 2003 (references^{83,84})

^b Those 39/95 with good compliance after one year were analyzed

^c weight loss was used as an intervention; not accounted for in the analyses; we expect a correlation anyway, because of regression to the mean.

^d 30 “randomly recruited from the Oxford Sleep clinic”, and the remaining 11 were oversampled on the basis of >4% dip in arterial O₂ saturation

^e OSLER test is an objective test like the maintenance of wakefulness test; assessed time until you start failing to respond to a blinking light

Note that all assessed outcomes have at best unclear clinical significance. All AHI measurements were performed with facility-based PSG. Ranges for the AHI at baseline were not reported in these studies. Studies are ordered by decreasing number of analyzed people.

Synopsis for section B1

Baseline AHI from facility-based PSG is only modestly associated with response to CPAP (as conveyed by the assessed outcomes) among people with *high probability for severe OSAHS (high AHI values on average)*. Thus, differences in baseline AHI cannot be used to accurately predict CPAP use or response to CPAP in *this population*. The same was true when associations with other indices (apart from AHI) were assessed. Therefore, there are indications that the exact value of AHI at baseline is not necessary for the prediction of CPAP response or use among people with high probability for OSAHS. This is in accordance with the conclusion of section A1. We will refer to this conclusion again in section B2 to explain the apparent “discrepancy” in the results of difference versus average analyses and sensitivity/specificity analyses.

We note that these results pertain to patients who had already been preselected for CPAP treatment. Such patients typically had very high baseline AHI on average and (very likely) symptomatic disease. Therefore these results cannot be extrapolated to

answer the question of whether laboratory-based PSG is useful in the management of people who are suspected for OSAHS.

In addition, a randomized study⁷³ suggested that the increased accuracy in AHI estimation with facility-based PSG versus a type IV monitor does not translate to a more accurate prediction of response to CPAP with respect to a quality of life outcome. This finding is in line with the aforementioned caveats and pertains to patients who are at the severe end of the AHI spectrum.

B2. How does the performance of portable monitors compare with facility-based polysomnography for the diagnosis of OSAHS?

We identified 75 eligible studies in total. Results are presented separately per type of monitor and per setting. The presentation follows the same pattern: A description of the studies precedes the presentation of findings on the concordance of individual measurements (difference versus average analyses). As discussed in section A2, these analyses answer the question of whether the compared monitors agree in the individual AHI (or RDI) measurements. Then we present analyses on the ability of portable monitors to detect AHI index suggestive of OSAHS with facility-based PSG (sensitivity/specificity analyses). These answer the question of whether the portable monitors are able to classify people similarly with facility-based PSG, irrespective of potential large differences in large AHI values. The synopsis section discusses the interpretation.

Note that some publications may be applicable to more than one sleep monitor type, or to more than one setting.

Type II monitors

Description of studies

We identified five studies⁸⁵⁻⁸⁹ that assessed the performance of type II monitors (**Table 7**). Two of these studies^{85,86} compared the same monitor in different settings (lab setting and home setting). As discussed in the Methods, these two studies were added in this section using a “best additional evidence approach” (because type II monitors are stripped down versions of comprehensive laboratory-based PSG). The number of analyzed patients in the five studies ranged from 20 to 99. The three studies that used type II monitors in the home setting^{85,86,89} were graded “B” for their overall quality, whereas the other two received grade “C”.

In all studies but one (Iber 2004), the assessed population had a high probability for OSAHS; the average AHI in facility-based polysomnography was 25 events/hour or more. Participants were referral cases in at least three out of five studies.^{85,87,89} Mean participant age was around 50 years, and in all studies the majority were males. Two studies^{85,88} used only automated scoring for the portable monitor, and one of them⁸⁵ used automated scoring for the hospital-based study as well. Patient hook-up was performed by trained technologists in all studies, and only in the smallest one the portable monitor was attended in a random half of the patients.⁸⁷

Table 7: Description of studies comparing type II monitors with facility-based PSG.

Author, Year, Country	Participants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/m ²)	Respiratory event definition for facility-based PSG and portable monitor	Name	Type II monitor Scoring	Att	Hook-up	Timing	Quality
<i>Portable monitor assessed at home</i>													
Gagnadoux, 2002 France	Referrals	111 (99)	55	29 (ND)	83	27.5	Apnea: No airflow (therm) Hypopnea: ↓Effort (Th/AB belts) ≥50%	Minisomno	Auto	no	Tech	ND	B
Portier, 2000 France	Referrals	103 (78)	52	26 (ND)	82	31.0	Apnea: ↓Airflow (therm) >75% Hypopnea: ↓Airflow (therm) >25-75%	Minisomno	Auto + manual edit	no	Tech	Δt=14d	B
Iber, 2004 US	Incident patients	76 (64)	53	10 (4, 23)	53	31.0	Apnea: No (or “almost no”) airflow (therm) Hypopnea: ↓Airflow (therm) “discernible” or ↓Effort (Th/AB belts) ≥30% Events associated with ↓3% SaO ₂ ^a	Compumedics PS-2	Manual	no	Tech	Δt=14d	B
<i>Portable monitor assessed in the laboratory</i>													
Orr, 1994 USA	ND	40 (40)	ND	26 (ND)	ND	ND	ND	Sleep I/T	Auto	ND	Tech	Simultaneous	C
Mykityn, 1999 Australia	Referrals	20 (20)	50	25 (1, 79) ^b	100	30.6	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% or ↓Effort (Th/AB belts)	Compumedics PS-1	Manual	50% yes ^c	Tech	Simultaneous	C

Att: attended; Auto: automated scoring AHI: apnea-hypopnea index in events/hour of sleep; d: days N_E/N_A: Number enrolled/analyzed; ND: not described; PSG: polysomnography; Tech: technologist; Th/AB belts: thoracoabdominal belts; therm: Thermistors used in airflow estimation; y: year(s); Δt: time interval between the two studies.

Respiratory events across all studies were of at least 10 seconds duration. As mentioned in the header row, respiratory events were defined identically in for the portable monitors as with laboratory-based PSG. Studies are ordered per setting and then by decreasing number of analyzed people.

^a set of criteria selected by the EPC for presentation. Other set of criteria (association with 4% desaturations, or irrespectively of desaturations yielded similar inferences)

^b Data from digitized graph

^c 50% were attended and 50% unattended, allocation was random.

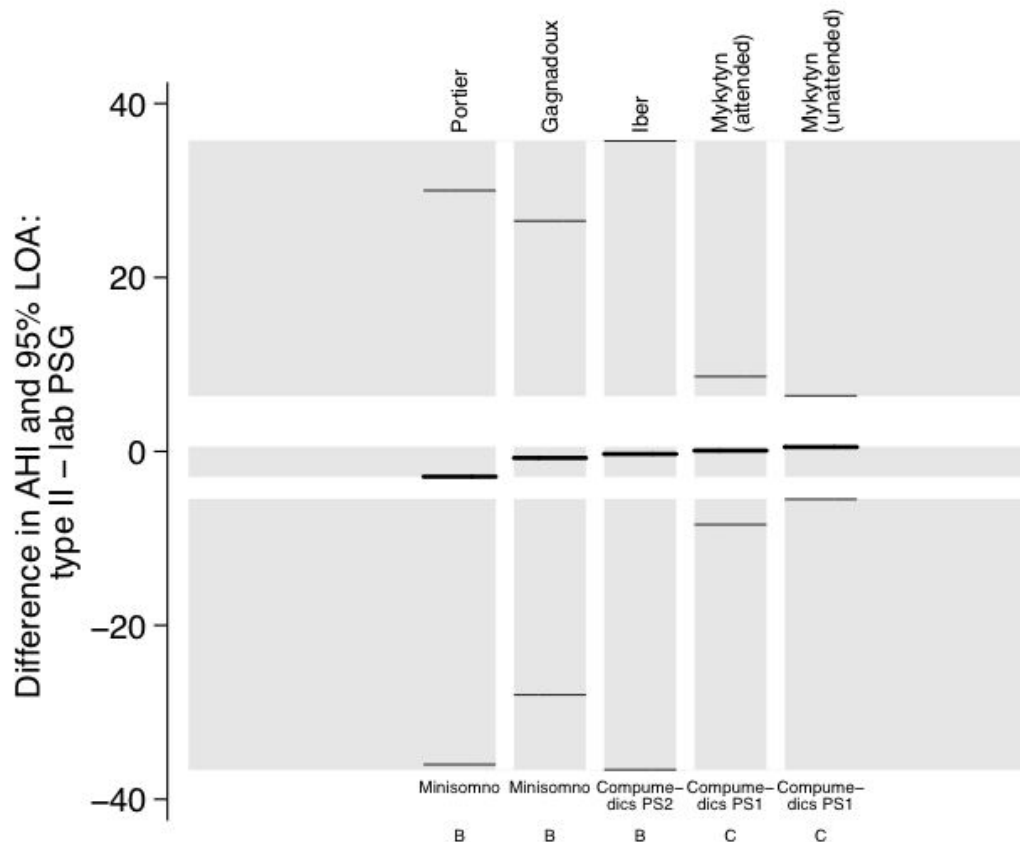
Findings-concordance

All studies, except for Orr 1994, assessed the agreement of facility-based PSG with type II monitors using difference versus average plots, or provided good quality graphs that allowed our EPC to perform these analyses.^{87,89} **Figure 6** gives an overall representation of the mean difference between the two measurements and the limits of agreement from each study. In Iber 2004,⁸⁶ the difference between the two measurements was dependent on their average value (portable overestimated lab-based measurements for AHI<20 events per hour, and underestimated it in more severe cases; however, measurements were not performed simultaneously). Only Mykytyn 1999⁸⁷ performed simultaneous measurements, and reported the tightest limits of agreement (from -6.9 to 7.5 events/hour). The limits of agreement in Mykytyn 1999⁸⁷ were similarly narrow when attended and unattended sleep studies were considered separately (from -8.4 to 8.6, and from -5.5 to 6.4 events/hour respectively; EPC analyses from digitized graphs).

As shown in **Figure 6**, discrepancies in the individual measurements may be substantial. In other words, if the exact AHI value were of interest across the whole spectrum of observed AHI values, type II monitors could not be used interchangeably with PSG. However, as discussed in section A1, for AHI values that are large, the exact values are not very useful. The synopsis of this section provides relative discussion.

Iber 2004 found a high intraclass correlation coefficient for the paired measurements (ICC=0.77).⁸⁶ In the same study, the weighted κ for agreement in classification of cases to the quartiles of AHI according to lab-based measurements was 0.57. Note that a high intraclass correlation coefficient is expected among methods that measure the same quantity and does not exclude clinically important differences between the two methods (**Figure 6**).

Figure 6. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type II monitors.



Schematic representation of the agreement between type II monitors and facility-based PSG as conveyed by difference versus average analyses. Each study is represented by three lines; these stand for the mean bias, and 95% limits of agreement from the difference versus average analyses. The upper, middle and lower grey areas group the upper 95% limits of agreement, the mean difference, and the lower 95% limits of agreement, respectively. Studies are ordered from left to right based on their setting, and by increasing bias. The monitor make and the overall methodological quality are also depicted. Note that the narrowest limits of agreement are for the attended and unattended strata of the only study that assessed a type II monitor simultaneously with facility-based PSG (Mykytyn). AHI is measured as events/hour of actual sleep.

Findings-ability to detect AHI suggestive of OSAHS

Three studies (Portier 2000, Mykytyn 1999 and Orr 1994) assessed the ability of a type II monitor to predict an AHI>15 events/hour with facility-based PSG. Sensitivity and specificity were very high in all studies (**Table 8**). Mykytyn 1999 and Orr 1994 had both a positive likelihood ratio>10 and a negative likelihood ratio of <0.1. Portier 2000 had a positive likelihood ratio>10. However, the total number of evaluated people was only 138 (**Table 8**).

Table 8. Sensitivity and specificity of type II monitors in the sleep laboratory setting to predict AHI>15 events/hour with laboratory-based PSG.

Author, Year Country	N _E (N _A)	AHI cutoff for type II monitor	AHI>15 in facility-based PSG (%)	Sensitivity (95% CI)	Specificity (95% CI)	Quality
Portier, 2000 France	103 (78)	15	47	81 (65, 92)	97 (83, 100)	B
Orr, 1994 USA	40 (40)	15	35	100 (59, 100)	100 (75, 100)	C
Myktyyn, 1999 Australia	20 (20)	15	63	100 (86, 100)	93 (68, 100)	C

AHI: apnea-hypopnea index; CI: confidence interval; N_A/N_E: number analyzed/enrolled.

For Myktyyn 1999, sensitivity and specificity were 100% among attended only or unattended only studies, as well. However, numbers are very small for meaningful inferences.

AHI is measured in events/hour.

Type III monitors

Overall description of studies using type III monitors

We identified 22 studies^{31,32,50,52-57,90-102} that compared type III monitors with facility-based PSG in various settings; their description and findings are shown in **Table 9** and **Table 10**. Ten papers described sleep studies with portable monitors that were performed in settings other than a specialized sleep facility^{31,32,53,54,90,93,96,97,100,101} (eight in the home setting^{31,32,54,90,96,97,100,101} and two in the hospital, but not in a sleep clinic or sleep laboratory^{53,93}) (**Table 9**). The remaining 15 papers described sleep studies with portable monitors that were performed in sleep laboratories, sleep clinic, or specialized sleep units in general^{31,32,50,52,54-57,91,92,94,95,98,99,102} (**Table 10**). At least 15 different type III monitors were used across all studies (two papers did not report the make of the portable monitor they used; **Table 9**).

The number of analyzed participants in these studies ranged from 20 to 116. The median number of subjects per study was 51 (interquartile range: 36, 68). Three studies were graded “A” for their overall methodologic quality,^{54,55,57} nine were graded “B”,^{32,50,53,91,94,96-99} and the remaining were graded “C” (**Table 9** and **Table 10**).

More on studies in the home setting or in the hospital setting (not sleep units)

For the ten studies that assessed portable monitors in settings other than specialized sleep units, the average participant age was 51 years on median (interquartile range: 50, 52). In all studies the majority of the participants were males, and in seven studies males comprised more than three fourths of the population (**Table 9**). Participants were referral cases for the evaluation of suspected sleep apnea and were recruited from sleep laboratories or sleep centers in seven studies, and from sleep clinics or other clinics in three studies. We note that this information cannot be used as a robust or valid proxy of the participants’ prior probability for OSAHS.

The time interval between the measurements with the portable monitor and facility-based PSG was less than 2 weeks in three studies, and less than 40 days in all but one study (in the latter case it ranged between 2 and 93 days,¹⁰¹ without any active treatment in the meanwhile). The order of measurements was randomly allocated in five cases. All type III monitor sessions in **Table 9** were classified as unattended. In one paper, a monitor was used that alerted the patient with recorded signals when mistakes in probe connection occurred or when probes were dislodged.³² The technologist received

feedback (channel recordings) from the portable monitor at least every 30 minutes in White 1995³¹ and the type III monitor was only “partially attended” in a hospital-based study (hospital, not a specialized sleep unit).⁹³ A technologist hooked-up the portable monitor probes (in all participants or in a proportion of them) in at least six studies.

Table 9. Description of studies comparing type III monitors with facility-based PSG in settings other than a specialized sleep clinic or unit.

Author, Year Country	Partici- pants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/ m ²)	Respiratory event definition for facility- based PSG	Name	Type III monitor Scoring: Respiratory event definition	Att	Hook- up	Ti- ming	Quality
<i>Portable monitor assessed at home</i>													
Dingli, 2003 UK	Referrals to sleep center	61 (50)	50	29 (ND)	77	31.0	Apnea: No airflow (cannula, therm) Hypopnea: ↓Effort (Th/AB) ≥ 50%	Embletta	Auto: Default settings Manual: Apnea: No airflow (cannula) Hypopnea: ↓Effort (Th/AB) ≥ 50%	No	P	Δt≤40d ^a	A
Quintana- Gallego, 2004 Spain	Out- Patients, cardio- logy clinic	90 (68)	56	12 (0, 62)	87	28.6	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥4% or arousal	Apno- screen II	Manual: Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥4%	No	Tech	Δt≤30d ^a	B
Reichert 2003, USA	Referrals to sleep center	51 (45)	52	29 (0, 123) ^c	74	30.0	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥2%	Nova- Som QSG	Auto: Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥2%	No ^b	P	Δt≤7d	B
Parra, 1997 Spain	Referrals to pneumo- logy clinic	89 (89)	54	34 (ND)	82	29.0	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) discernible with paradoxical motion (Th/AB) or cyclical dip in SaO ₂	Eden- Trace	Manual: Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) discernible with paradoxical motion (Th/AB) or cyclical dip in SaO ₂	No	Tech (n=50) P (n=39)	Δt≤30d ^a	B
Whittle, 1997 UK	Referrals to sleep clinic	23 ^d (20)	50	31 (2, 67) ^c	83	ND	Apnea: No airflow (therm) Hypopnea: ↓Effort (Th/AB) ≥ 50%	Eden- Trace	Manual: Apnea: No airflow (therm) Hypopnea: ↓Effort (Th/AB) ≥ 50%	No	ND	ND	C
Ancoli-Israel, 1997 US	ND ^e	36 (34)	49	42 (2, 170)	94	ND	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50%	Night- watch	Auto + manual edit: Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50%	No	Tech	ND	C
White, 1995 USA	Referrals to sleep centers	72 (70)	48	28 (0, 133) ^c	74	33.0	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥4% or EEG arousal	Night- Watch	Auto + manual edit: Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥4% or arousal (leg movements)	No ^f	Tech	Δt=10d	C
Yin M, 2006 Japan	Referrals	44 (44)	52	34 (1, 88) ^c	90	27.0	Apnea: "Published criteria" Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥3%	Stardust II	Auto: Apnea: "Published criteria" Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥3%	No	P	Δt= 2-93d ^g	C

Author, Year Country	Partici- pants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/ m ²)	Respiratory event definition for facility- based PSG	Name	Type III monitor Scoring: Respiratory event definition	Att	Hook- up	Ti- ming	Quality
<i>Portable monitor assessed in hospital but not in a specialized sleep facility</i>													
Carasco, 1996 Spain	Referrals to sleep lab	36 (36)	52	35 (ND)	81	32	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) discernible with arousal or cyclical dip in SaO ₂	NS	Auto: Apnea: ↓Airflow ≥80% Hypopnea: ↓Airflow ≥ 35% with ↓SaO ₂ ≥2% Manual: Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) discernible or cyclical dip in SaO ₂	No	ND	Δt≤14d ^a	B
Lloberes, 1996 Spain	Referrals to sleep clinic	76 (76)	51	32 (ND)	71	31	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) discernible with arousal or cyclical dip in SaO ₂	NS	Auto: Apnea: ↓Airflow ≥80% with ↓SaO ₂ ≥2% for ≥30s Hypopnea: ↓Airflow ≥50% and ≤80% with ↓SaO ₂ >2% for ≥30s	No ^h	Tech	Δt≤21d ^a	C

Att: attended; Auto: automated scoring AHI: apnea-hypopnea index in events/hour of sleep; d: days; EEG: electroencephalogram; N_A/N_E: Number analyzed/enrolled; ND: not described; P: study participant; PSG: polysomnography; Tech: technologist; Th/AB belts: thoracoabdominal belts; therm: Thermistors used in airflow estimation; y: year(s); Δt: time interval between the two studies.

Respiratory events across all studies were of at least 10 seconds duration unless otherwise noted. Studies grouped by setting and ordered by overall quality.

^a Random order of the two sleep studies.

^b Unattended, but the device has built in controls to check for probe failure/no signal, and wakes the patient up with recorded messages.

^c Data from digitized graph.

^d Only the validation study in this paper is eligible.

^e Study population comprises of participants in an epidemiologic study.

^f Data sent with a modem to the sleep center every 30 minutes; the technologist called the patient at home to correct the probes if needed.

^g No active treatment between the two studies.

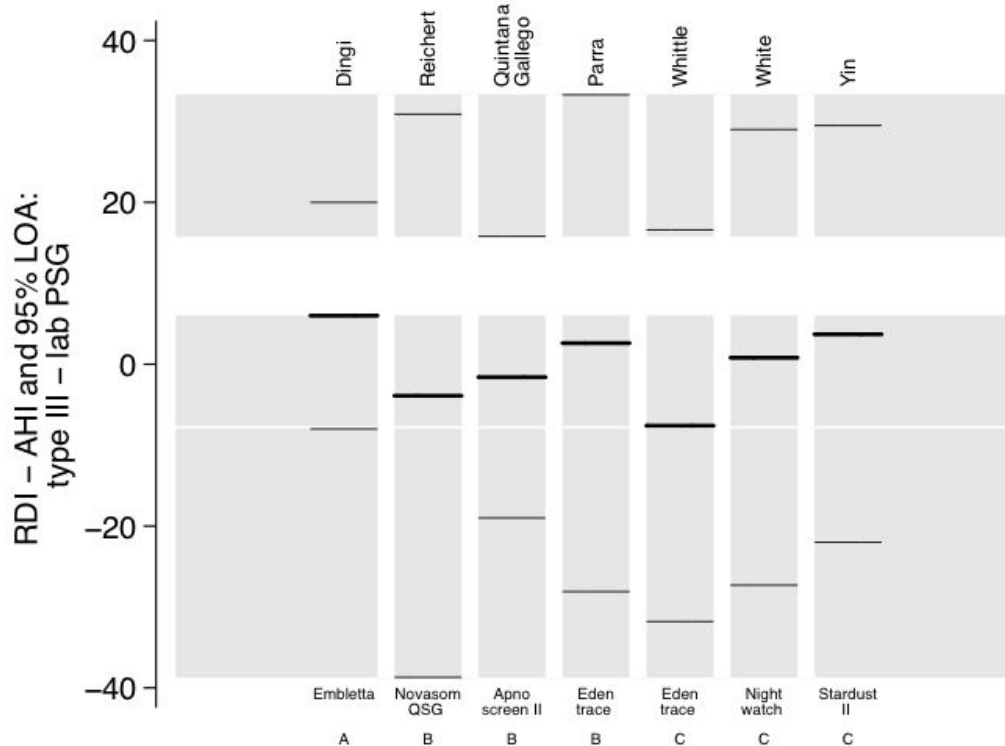
^h verbatim: "partially attended".

Findings of studies in the home setting - concordance

Seven of the eight studies in the home setting performed difference versus average analyses or provided enough information that allowed our EPC to perform such analyses.^{31,32,54,96,97,100,101} The Embletta, NovaSom QSG, EdenTrace, Apnoscreen II, Stardust II and NightWatch monitors were used in these seven studies (**Table 9**). Reichert 2003³² used data from 3 nights of recording with the portable monitor (NovaSom QSG). As evident in **Figure 7**, the 95% limits of agreement could not exclude quite large differences in the measurements with the two monitors (even differences as large as 20 or 35 events/hour, depending on the monitor). Moreover, in two studies^{54,100} the absolute mean difference of the measurements were more than 5 events per hour. The difference between the two measurements was statistically significantly dependent on their average in four studies^{32,97,100,101} (Bradley-Blackwood F test, $p < 0.05$; our analyses). This usually means that the absolute difference in the measurements is more pronounced for people with higher AHI. In these analyses all monitors were unattended, and most were manually scored (or at least used manual editing of automated scoring). Scoring was automated only in Yin 2006¹⁰¹ and Reichert 2003.³² Overall, the width of the limits of agreement did not change with the time interval between the two measurements.

Finally, two studies^{32,54} assessed agreement in classification of people above and below the threshold of 15 events/hour. Dingli 2003⁵⁴ found good agreement ($\kappa = 0.62$) with facility-based PSG when the Embletta monitor recordings were scored manually, and poor agreement ($\kappa = 0.10$) when automated scoring was used. Reichert 2003 found good agreement ($\kappa = 0.73$) with automated scoring of NovaSom QSG recordings.³²

Figure 7. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type III monitors in the home setting.

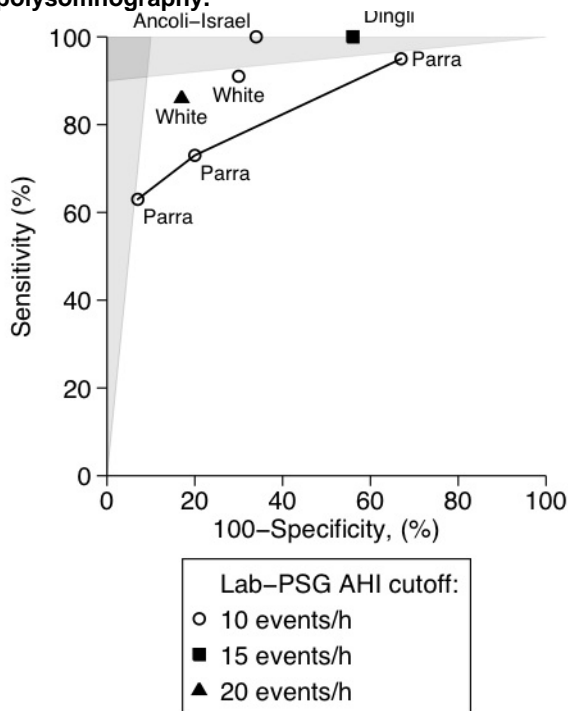


Schematic representation of the agreement between portable monitors and facility-based PSG as conveyed by difference versus average analyses. Each study is represented by three lines; these stand for the mean bias, and 95% limits of agreement from the difference versus average analyses. The upper, middle and lower grey areas group the upper 95% limits of agreement, the mean difference, and the lower 95% limits of agreement, respectively. Monitor make and overall methodologic quality are also depicted in the bottom of the graph. Studies are ordered from left to right based on their overall quality and by increasing bias. AHI is measured as events/hour of actual sleep.

Findings of studies in the home setting - predicting AHI suggestive of OSAHS

Four studies assessed the sensitivity and specificity of portable monitor recordings in the home setting to identify AHI suggestive of OSAHS.^{31,54,90,96} The cutoffs of AHI in facility-based PSG that were considered suggestive of OSAHS were 15 events/hour of sleep (one study⁵⁴), 10 events/hour of sleep (three studies^{31,90,96}) and 20 events/hour of sleep (one study³¹). Parra 1997 reported sensitivity and specificity pairs for three cutoffs of the RDI index derived from the type III monitor (8, 10 and 23 events/hour of recording). **Figure 8** illustrates the diagnostic ability of these four studies on a square plot. Only Dingli 2003 and Ancoli-Israel 1997 had a negative likelihood ratio of less than 0.1. None had a positive likelihood ratio of 10 or more. However, as evident from the graph, the remaining studies were also near the boundary of the regions defining high positive likelihood ratios or low negative likelihood ratios.

Figure 8. Diagnostic ability of type III monitors in the home setting to identify AHI suggestive of OSAHS in laboratory-based polysomnography.



Square plot (plot in the ROC space) depicting the five studies that estimated the sensitivity and specificity of type III monitors to predict an AHI suggestive of OSAHS. Sensitivity/specificity pairs from the same study (obtained with different cutoffs for the type III monitor) are connected with lines. These lines are not representative of the ROC curve of the pertinent studies. Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded area have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both $LR+ > 10$ and $LR- < 0.1$. AHI: Apnea-hypopnea index.

Findings of studies in the hospital setting (not sleep units) - concordance

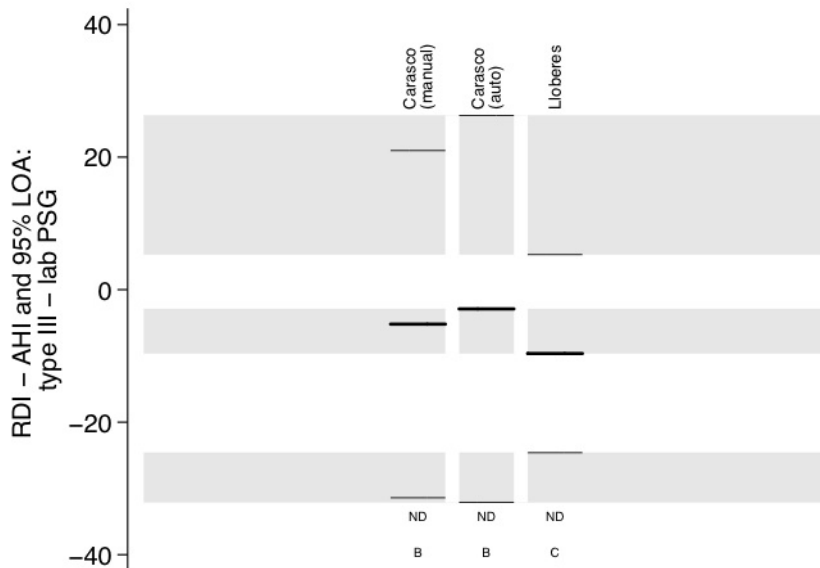
The two relevant studies performed difference versus average analyses or provided enough information that allowed our EPC to perform such analyses.^{53,95} (Table 9). Figure 9 illustrates the findings of difference versus average analyses. Carasco 1996⁵³ provided information both for automated and manual scoring of the portable monitor recordings.

The concordance of the individual measurements between the portable monitor and facility-based PSG was suboptimal. On average type III monitors underestimated facility-based PSG measurements by three to 11 events/hour. The spread of the 95% limits of agreement did not exclude an underestimation of the facility-based PSG AHI by 20 events/h with the portable monitors. (Figure 9) All studies were performed within 3 weeks from each other, and scoring methods for the portable monitor differed (automated in Lloberes 1996 and both methods in Carasco 1996). Any differences in concordance compared to the home setting should not be attributed to the setting per se. Differences in the operational characteristics of the monitors or the scoring algorithms may also play a role.

Finally, Carasco 1996 described good concordance between the two monitors in classifying subjects above or below an AHI (or RDI for the portable monitor) of 20 events/hour. Using manual, scoring κ was 0.77; and using the best-performing algorithm for automated scoring, it

was 0.60 (no statistical significance levels were provided). Note that the best-performing algorithm might not yield similarly high results in different subjects, due to overfitting (the algorithm might fit very well the recordings from the particular set of subjects, but may perform poorly in different subjects).

Figure 9. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type III monitors in the hospital setting (not specialized sleep units).



Schematic representation of the agreement between portable monitors and facility-based PSG as conveyed by difference versus average analyses. Each study is represented by three lines; these stand for the mean bias, and 95% limits of agreement from the difference versus average analyses. The upper, middle and lower grey areas group the upper 95% limits of agreement, the mean difference, and the lower 95% limits of agreement, respectively. Monitor make and overall methodologic quality are also depicted in the bottom of the graph. Studies are ordered from left to right based on their overall quality, and by increasing bias.

Note that for the automated scoring of Carasco, the algorithm with the best concordance with facility-based PSG is selected among many that were developed. Other algorithms had similar or much worse performance. AHI is measured as events/hour of actual sleep. ND: Not described (for the make of the portable monitor).

Findings of studies in the hospital setting (not sleep units) - predicting AHI suggestive of OSAHS

No data on the sensitivity or specificity of the portable monitors to predict AHI suggestive of OSAHS were reported in this setting.^{53,93}

More on studies in specialized sleep center settings

For the 15 studies of type III monitors in the laboratory setting^{31,32,50,52,54-57,91,92,94,95,98,99,102} the average participant age was 51 years on median (interquartile range: 47, 53) (**Table 10**). The majority of the patients were male (except for Su 2004,⁹⁸ – numbers based on studies that reported information on mean age). Participants were mostly referral cases for the evaluation of suspected sleep apnea. Sampling from the general population (in the context of an epidemiologic study) was used in Ballester 2000.⁹¹ We caution that the sampling population is not a perfect proxy for the participants' prior probability of OSAHS.

Three papers explicitly reported that the sleep monitor was attended by a technologist,^{32,50,94} and in one case the monitor reported the recordings to the technologist at

least every 30 minutes.³¹ It is likely that the technologist affixed the portable monitor's probes in all cases, but this was clearly stated in 9 papers (**Table 10**).

Table 10. Description of studies comparing type III monitors with facility-based polysomnography in the sleep clinic or sleep unit.

Author, Year Country	Participants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/m ²)	Respiratory event definition for facility-based PSG	Name	Type III monitor Scoring: Respiratory event definition	Att	Hook-up	Quality
Dingli, 2003 UK	Referrals to sleep center	61 (50)	50	29 (ND)	77	31.0	Apnea: No airflow (cannula, therm) Hypopnea: ↓Effort (Th/AB) ≥ 50%	Embletta	Auto: Default settings Manual: Apnea: No airflow (cannula) Hypopnea: ↓Effort (Th/AB) ≥ 50%	No	Tech	A
Ficker, 2001 Germany	Referrals to sleep center	51 (51)	53	24 (0, 111)	86	29.0	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥4%	Somno-check	Auto: [same as manual] Manual: Apnea: No airflow (therm) Hypopnea: ↓ airflow (therm) discernible with ↓SaO ₂ ≥4%	ND	Tech	A
Zucconi, 1996 Italy	Referrals to sleep center	30 (29)	53	32 (1, 86) ^a	68	30.7	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥40%	Micro Digi-traper-S	Auto: Apnea/Hypopnea: ↓Airflow-Effort (therm-Th/AB) ≥40% Auto + manual edit: As above with visual check Manual: Apnea: No airflow (therm) Hypopnea: ↓ airflow (therm) discernible with ↓SaO ₂ ≥3%	No	ND	A
Ballester, 2000 Spain	General population	116 (116)	47	10 (0, 84) ^a	56	26.0	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) discernible with arousal or ↓SaO ₂ ≥3%	Sibel Home-300	Manual: Apnea: No airflow (therm) Hypopnea: ↓ airflow (therm) discernible with ↓SaO ₂ ≥3%	ND	Tech	B
Man, 1995 Canada	Referrals to sleep clinic	104 (104)	47	17 (ND)	78	ND	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) >50%	PolyG	Manual: Apnea: No airflow (therm) Apnea/Hypopnea: ↓Airflow (therm) ≥50%	Yes	Tech	B
Su, 2004 USA	Referrals to sleep clinic	60 (60)	45	27 (2, 122) ^a	42	35.6	Apnea: No airflow (therm) with ↓Effort (Th/AB) ≥70% and ↓SaO ₂ ≥4% Hypopnea: No airflow (therm) with ↓Effort (Th/AB) ≥30% and ↓SaO ₂ ≥4%	SNAP	Auto + manual edit: Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) discernible with ↓SaO ₂ ≥4%	ND	Tech	B
Verse, 2000 Germany	Referrals to sleep clinic	53 (53)	48	18 (0, 76)	92	27.4	Apnea: ↓Airflow (therm) >80% Hypopnea: ↓Airflow (therm) >50%	Poly-Mesam	Auto: Apnea: ↓Airflow (therm) >80% Hypopnea: ↓Airflow (therm) >50% Manual: ND (same as auto?)	ND	Tech	B

Author, Year Country	Participants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/m ²)	Respiratory event definition for facility-based PSG	Name	Type III monitor Scoring: Respiratory event definition	Att	Hook-up	Quality
Reichert 2003, USA	Referrals to sleep center	51 (44)	52	29 (0, 123) ^a	74	30.0	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥2%	Nova-Som QSG	Auto: Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% with drop in SaO ₂ ≥2%	Yes	ND	B
Redline 1991, USA	Various	25 (25)	53	37 (0, 102)	ND	31.4	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) discernible with ↓SaO ₂ ≥4% or ↓SaO ₂ ≥2% and arousal	Edent 4700	Manual: Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) discernible with ↓SaO ₂ ≥4%	Yes ^b	ND	B
Calleja, 2002 Spain	Referrals to sleep lab	86 (79)	52	34 (ND)	89	30.1	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) discernible with arousal or ↓SaO ₂ ≥3%	Merlin	Manual: Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) discernible with ↓SaO ₂ ≥3%	No	Tech	C
Fietze, 2002 Germany	Referrals	66 (66)	51	24 (ND)	98	32.9	Apnea: ↓Airflow (therm) >85% Hypopnea: ↓Airflow (therm) >50% with ↓SaO ₂ ≥3% ^c	Merlin	Auto: ND Manual: Apnea: ↓Airflow (therm) >85% Hypopnea: ↓Airflow (therm) >50% with ↓SaO ₂ ≥3% ^c Auto: Apnea/Hypopnea: ↓SaO ₂ ≥3% ^c	ND	ND	C
Emsellem, 1990 US	Referral to sleep center	67 (63)	45	ND	ND	ND	Apnea/hypopnea: ↓Airflow (therm, CO ₂ gauge) ≥50%	Eden-Trace	Manual: Apnea/hypopnea: ↓airflow (therm) ≥50%	ND	ND	C
Marrone, 2001 Italy	Referrals to sleep lab	50 (50)	50	ND	80	ND	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) "noticeably" with drop in SaO ₂ ≥4%	Poly-Mesam	Auto + manual edit: Apnea: No airflow (therm) Hypopnea: ↓Airflow "noticeably" with drop in SaO ₂ ≥4%	No	Tech	C
White, 1995 USA	Referrals to sleep clinic	30 (30)	51	31 (0, 135) ^a	77	33	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥4% or EEG arousal	Night-Watch	Auto + manual edit: Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥4% or arousal (leg movements)	No ^d	Tech	C
Claman, 2001 USA	Referrals to sleep center	42 (42)	54	26 (0, 90) ^a	74	30.6	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥4%	Bedbugg	Auto: [not described]	ND	ND	C

Att: attended; Auto: automated scoring AHI: apnea-hypopnea index in events/hour of sleep; d: days; EEG: electroencephalogram; N_E/N_A: Number enrolled/analyzed; ND: not described; PSG: polysomnography; Tech: technologist; Th/AB belts: thoracoabdominal belts; therm: Thermistors used in airflow estimation; y: year(s); Δt: time interval between the two studies.

Respiratory events across all studies were of at least 10 seconds duration unless otherwise noted. Studies ordered by overall quality, and then by sample size after grouping per type III monitor make.

^a Data extracted from digitized graphs and corroboration with text.

^b in 20/25 people who were studied in the lab; 5/25 were studied at home (not separable).

^c In addition to apnea and hypopnea, periodic breathing was considered: ↓Airflow (therm) >50% with ↓SaO₂≥2% (irrespectively of duration).

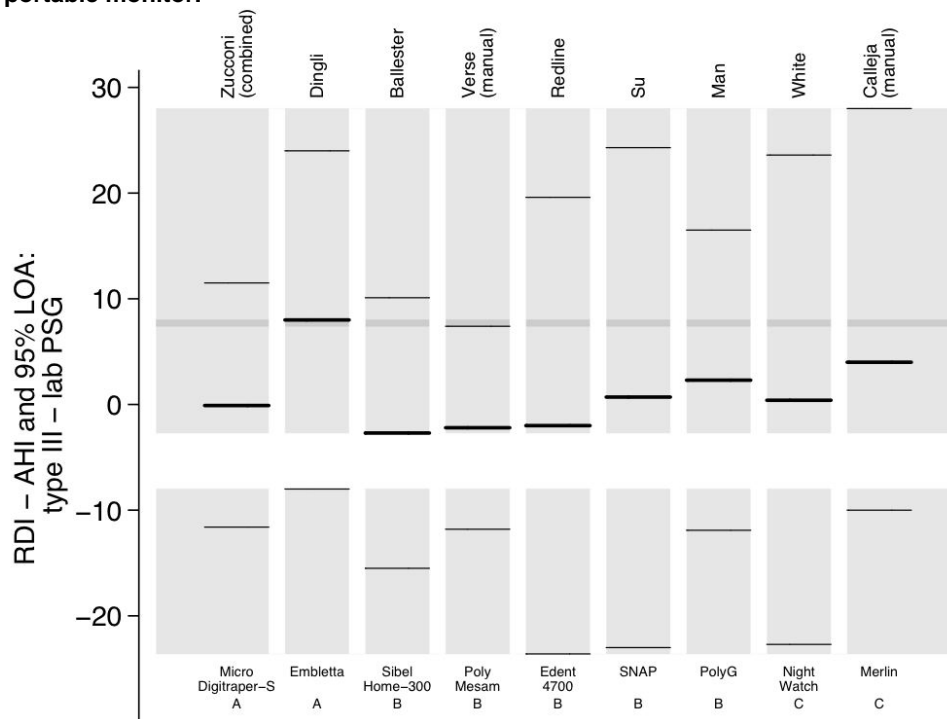
^d Unattended, but data were downloaded at least every 30 minutes and the technologist had the ability for an overview.

Findings of studies in the specialized sleep unit setting – concordance

The concordance of the individual measurements between type III monitors and facility-based PSG was assessed with difference versus average plots in several studies. To enhance clarity in the presentation these analyses are depicted in separate figures (**Figure 10**, **Figure 11** and **Figure 12**) for various sets of studies. In almost all studies the difference of the measurements was dependent on their average, so the 95% limits of agreement should be interpreted with caution.

Figure 10 illustrates difference versus average analyses from studies that used manual scoring or automated scoring with manual editing.^{31,50,52,54,57,91,94,98,99} The differences in the measurements varied across monitor makes. However, monitor make may not be the only explanation for the variability. Again, the 95% limits of agreement show suboptimal agreement in the measurements (see the synopsis of Section B2 for an interpretation of this finding).

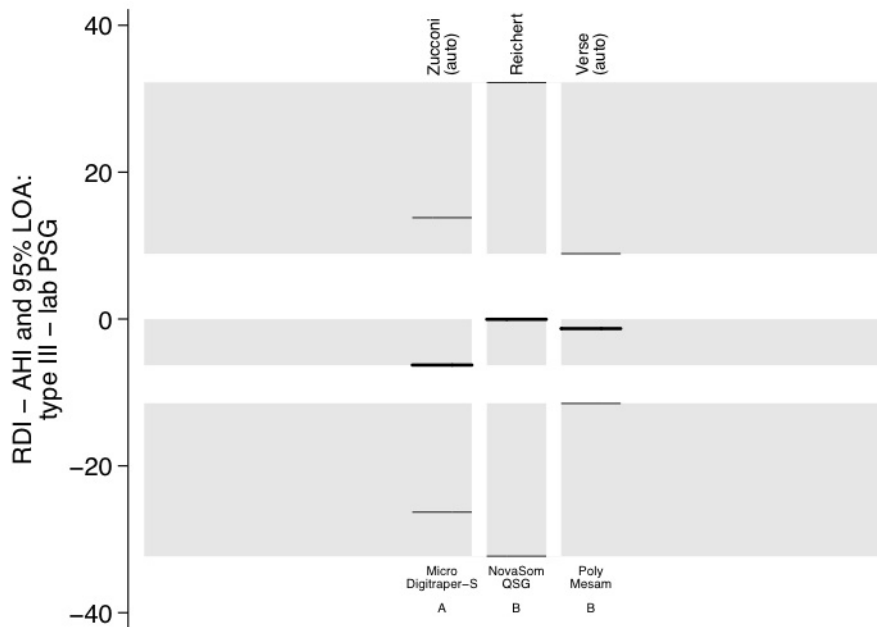
Figure 10. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type III monitors in specialized sleep units. Studies that used manual scoring for the portable monitor.



Schematic representation of the agreement between portable monitors and facility-based PSG as conveyed by difference versus average analyses. Each study is represented by three lines; these stand for the mean bias, and 95% limits of agreement from the difference versus average analyses. The upper, middle and lower grey areas group the upper 95% limits of agreement, the mean difference, and the lower 95% limits of agreement, respectively. Note that the upper and middle grey areas overlap slightly. The make of the monitor and the overall study quality are also depicted in the lower part of the graph. Only studies that used both apneas and hypopneas in the definition of respiratory events for both monitors are shown.

Figure 11 illustrates difference versus average analyses from studies that used automated scoring and were rated grade “A” or “B” for their overall methodologic quality.^{32,57,99} As was the case for manual scoring, the differences in the measurements varied across monitor makes. As evident from the graph, the 95% limits of agreement show suboptimal agreement in the measurements (see the synopsis of Section B2 for an interpretation of this finding).

Figure 11. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type III monitors in specialized sleep units. Good and moderate quality studies that used automated scoring for the portable monitor.

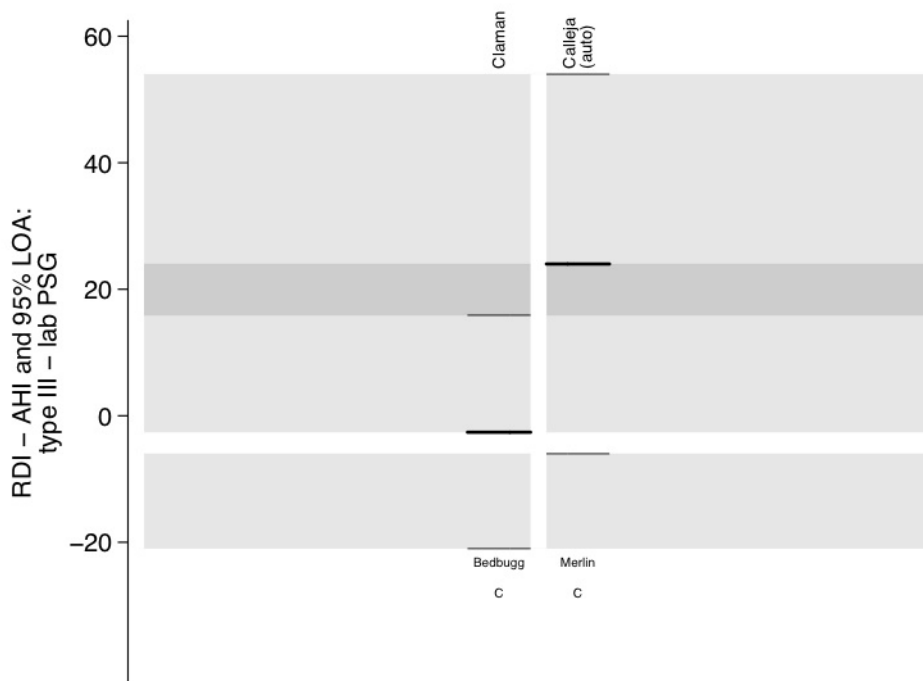


Schematic representation of the agreement between portable monitors and facility-based PSG as conveyed by difference versus average analyses. Each study is represented by three lines; these stand for the mean bias, and 95% limits of agreement from the difference versus average analyses. The upper, middle and lower grey areas group the upper 95% limits of agreement, the mean difference, and the lower 95% limits of agreement, respectively. The make of the monitor and the overall study quality are also depicted in the lower part of the graph.

Only studies of good and moderate quality that used both apneas and hypopneas in the definition of respiratory events for both monitors are shown.

Figure 12 illustrates difference versus average analyses from studies that used automated scoring and were rated grade “C” for their overall methodologic quality. There was extreme variability in the mean bias and 95% limits of agreement obtained from these studies. As evident from the graph, the 95% limits of agreement show suboptimal agreement in the measurements (see the synopsis of Section B2 for an interpretation of this finding).

Figure 12. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type III monitors in specialized sleep units. Poor quality studies that used automated scoring for the portable monitor.



Schematic representation of the agreement between portable monitors and facility-based PSG as conveyed by difference versus average analyses. Each study is represented by three lines; these stand for the mean bias, and 95% limits of agreement from the difference versus average analyses. The upper, middle and lower grey areas group the upper 95% limits of agreement, the mean difference, and the lower 95% limits of agreement, respectively. Note that the upper and middle grey areas overlap extensively. The make of the monitor and the overall study quality are also depicted in the lower part of the graph. Only studies of poor methodological quality that used both apneas and hypopneas in the definition of respiratory events for both monitors are shown.

Reichert 2003 calculated high agreement ($\kappa=0.86$) between the NovaSom QSG monitor and facility-based PSG in the classification of patients above or below a cutoff of 15 events/hour in the AHI. For the same AHI threshold, Dingli 2003⁵⁴ found good agreement ($\kappa=0.50$) with facility-based PSG when the Embletta monitor recordings were scored manually, and poor agreement ($\kappa=0.28$) when automated scoring was used.

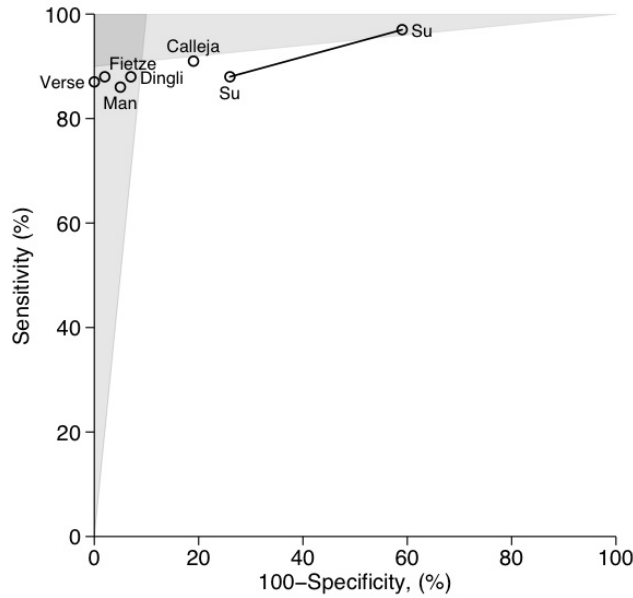
Findings of studies in the specialized sleep unit setting - predicting AHI suggestive of OSAHS

Fourteen studies^{31,32,50,52,54-57,91,94,95,98,99,102} assessed the ability of type III monitors in the laboratory setting to predict AHI in facility-based polysomnography that were suggestive of OSAHS (using a cutoff of 15 events/hour of sleep,^{32,52,54,56,94,98,99,102} 10 events/hour of sleep^{31,50,52,55-57,91,95,102} and 20 events/hour of sleep^{31,52,55,102}). Other cutoffs for AHI (5, 30 or 40 events/hour were also used in some of the aforementioned studies and the remaining study⁹²).

Figure 13 shows studies that used manual scoring or automated scoring with manual editing for type III monitors, and employed the cutoff of 15 events/hour in AHI as

suggestive of OSAHS. All studies lie in or close to the regions corresponding to high positive likelihood ratio or low negative likelihood ratio.

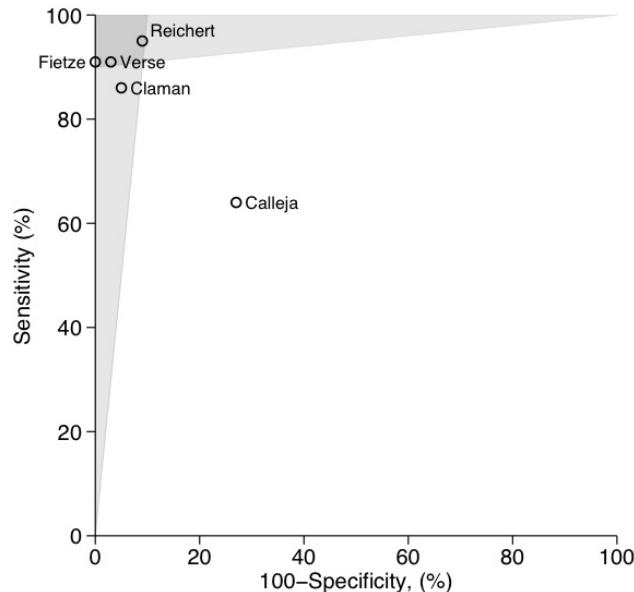
Figure 13. Diagnostic ability of type III monitors in specialized sleep units to identify AHI>15 events/hour in laboratory-based polysomnography. Manual or combined manual and automated scoring.



Sensitivity/specificity pairs from the same study (obtained with different cutoffs for the type III monitor) are connected with lines. These lines are not representative of the ROC curve of the pertinent studies. Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded area have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both $LR+ > 10$ and $LR- < 0.1$. The figure depicts studies that used manual scoring or combined manual and automated scoring for the type III monitor, and a cutoff of 15 events/h as suggestive of OSAHS in facility-based PSG.

Figure 14 shows studies that used automated scoring for the type III monitor for the same cutoff of 15 events/hour in facility-based PSG. Almost all studies lie in or very near the regions corresponding to high positive likelihood ratio or low negative likelihood ratio.

Figure 14. Diagnostic ability of type III monitors in specialized sleep units to identify AHI>15 events/hour in laboratory-based polysomnography. Automated scoring.

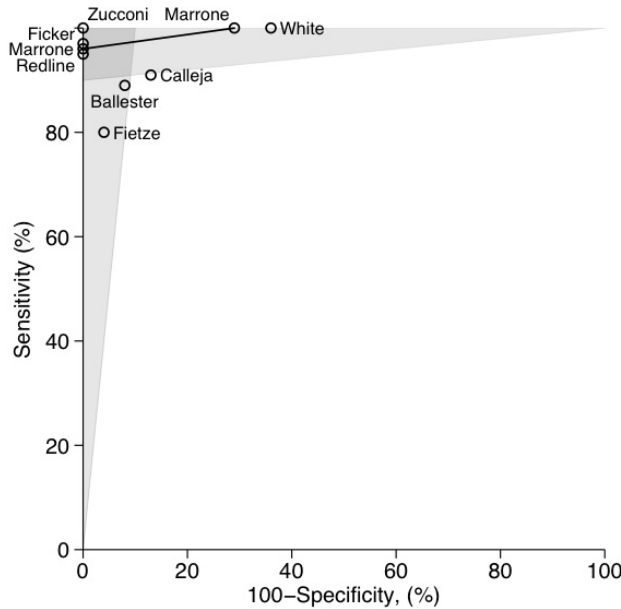


Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded area have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both $LR+ > 10$ and $LR- < 0.1$. The figure depicts studies that used automated scoring for the type III monitor, and a cutoff of 15 events/h as suggestive of OSAHS in facility-based PSG.

When a cutoff of 10 events/hour was employed in facility-based polysomnography as suggestive of OSAHS (**Figure 15**), studies of type III monitors using manual or combined manual and automated scoring had good predictive ability. Four studies had very high positive likelihood ratio and very low negative likelihood ratio.

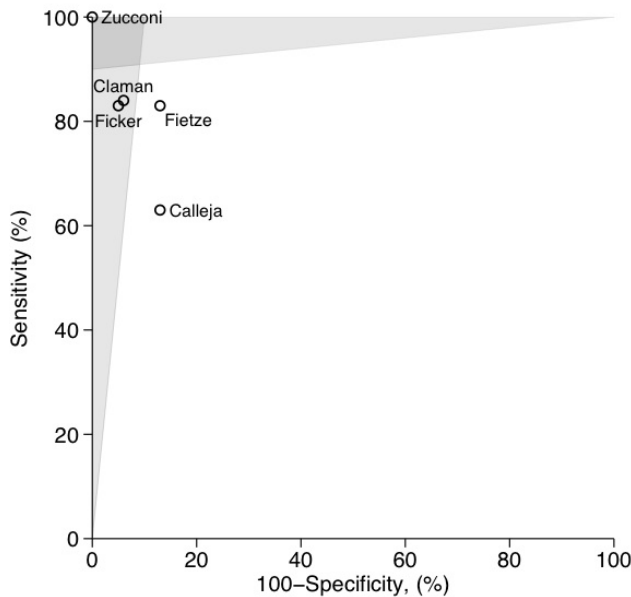
The scatter of the studies was greater when automated scoring was used instead of manual scoring (**Figure 16**). Only one study had very high positive likelihood ratio and very low negative likelihood ratio for automated scoring (**Figure 16**).

Figure 15. Diagnostic ability of type III monitors in specialized sleep units to identify AHI>10 events/hour in laboratory-based polysomnography. Manual or combined manual and automated scoring.



Sensitivity/specificity pairs from the same study (obtained with different cutoffs for the type III monitor) are connected with lines. These lines are not representative of the ROC curve of the pertinent studies. Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded area have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both $LR+ > 10$ and $LR- < 0.1$. The figure depicts studies that used manual scoring or combined manual and automated scoring for the type III monitor, and a cutoff of 10 events/h as suggestive of OSAHS in facility-based PSG.

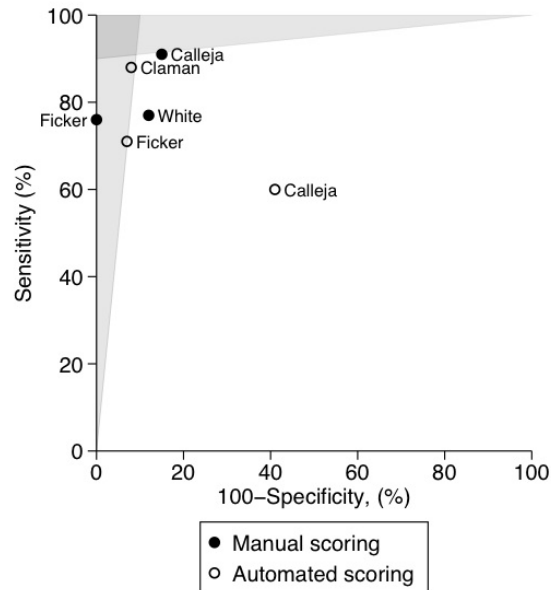
Figure 16. Diagnostic ability of type III monitors in specialized sleep units to identify AHI>10 events/hour in laboratory-based polysomnography. Automated scoring.



Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded area have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both $LR+ > 10$ and $LR- < 0.1$. The figure depicts studies that used automated scoring for the type III monitor, and a cutoff of 10 events/h as suggestive of OSAHS in facility-based PSG.

For the cutoff of 20 events/hour of sleep, manual scoring for type III monitors yielded comparatively worse diagnostic ability (**Figure 17**). Automated scoring for this cutoff resulted in quite different diagnostic performances across the various studies (**Figure 17**). The differential scatter of study points on the plot was not readily explained by the study characteristics that have been assessed in this technology assessment.

Figure 17. Diagnostic ability of type III monitors in specialized sleep units to identify AHI>20 events/hour in laboratory-based polysomnography.



Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded area have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both $LR+ > 10$ and $LR- < 0.1$. The figure depicts studies that used either manual scoring or automated scoring for the type III monitor, and a cutoff of 20 events/h as suggestive of OSAHS in facility-based PSG.

Type IV monitors

Overall description of studies using type IV monitors

We identified 46 studies that compared type IV monitors with facility-based PSG in various settings.^{26-30,42,58,59,76,80-84,103-134} As mentioned before, in this technology assessment any monitor that did not fulfill the criteria for type III was considered to be a type IV monitor. We also include a study of SNAP, a type III monitor, where only a subset of the available channels was utilized (and therefore the type of the monitor was downgraded).¹¹⁶ Eleven studies that were conducted in the home setting^{42,59,81,105,111,118,122-124,127,128} (as well as one study conducted in the hospital but not in a specialized sleep facility⁷⁶ and a study in which the setting is unclear¹⁰⁶) are described in **Table 11** (monitors assessing more three or more bioparameters) and **Table 12** (monitors assessing two or less bioparameters). The 38 studies that were conducted in specialized sleep centers^{26-30,42,58,59,80-84,103-105,107-110,112-117,119-121,125-127,129-134} are reported in **Table 13**

(monitors assessing more three or more bioparameters) and **Table 14** (monitors assessing two or less bioparameters).

Across the 46 studies, at least 11 different makes of type IV monitor were used (apart from the use of oximeters as portable monitors). Overall, the median number of subjects analyzed was 63 (interquartile range: 34, 140). Only Westbrook 2005¹²⁷ and Ayappa 2004⁷⁶ were graded “A” for overall methodological quality. Twenty one studies received grade “B”,^{28,30,59,80,83,84,108,111-116,118,123,124,126,129-132} and the remaining received grade “C”.

More on type IV monitors in the home setting

For the eleven studies that assessed type IV monitors in the home setting (and the remaining two studies – one conducted in the hospital in a non-specialized facility and one study where their setting was unclear) the average participant age was 51 years on median (interquartile range: 46, 55). The median number of subjects analyzed was 74 (interquartile range: 40, 114). Over 70% of the participants were males in the ten studies that reported gender distributions (**Table 11** and **Table 12**). Subjects in all studies were referral cases to sleep laboratories or specialized sleep centers, and none of the studies sampled from the general population.

The time interval between the measurements with the portable monitor and facility-based PSG was less than a month in all studies that reported this information, and less than a week in three of them. None reported that the order of measurements with the two methods (type IV monitors or facility-based PSG) was random. Probes were affixed by a technologist in two studies,^{76,123} by the subjects themselves after brief instructions in four studies,^{42,81,105,124} or randomly by technologist or the subjects themselves in Golpe 2002¹¹¹ and Pang 2006.¹¹⁸ No information was given for the remaining five studies.

As mentioned in the Methods, we present separately findings for type IV monitors according to the number of bioparameters they recorded.

Table 11: Description of studies comparing type IV monitors (three or more bioparameters) with facility-based polysomnography in settings other than a specialized clinic or unit.

Author, Year Country	Parti-cipants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/m ²)	Respiratory event definition for facility-based PSG	Name [Signals]	Portable Scoring: Respiratory event definition	Att	Hook-up	Ti-ming	Quality
<i>Portable monitor assessed at home</i>													
Westbrook, 2005 USA	Referrals from sleep centers (mostly)	191 (187)	46	27 (0, 118) ^a	66	ND	ND	ARES [SaO ₂ ; head position; snoring]	Auto: ↓SaO ₂ >2.2%, or ↓SaO ₂ ≤2.2% with resaturation at 2.2% with arousal (based on activity) ^b Manual: ↓SaO ₂ ≥ 4%, or ↓SaO ₂ ≥ 2% with visible change in HR	ND	ND	ND ^c	A
Schafer, 1997 Germany	Referral patients	114 (114)	56	29 (ND)	88	30.8	Apnea: No airflow (therm). Hypopnea: ↓Airflow (therm) ≥ 50% ↓SaO ₂ ≥ 4% or arousal	Mesam IV [SaO ₂ ; body position; HR; snoring]	Auto: ↓SaO ₂ ≥ 4%, or ↓SaO ₂ ≥ 2% with visible change in HR	ND	Tech	ND	B
Golpe, 2002 Spain	Referrals to a sleep center	55 (44)	53	ND	96	30.3	Apnea: No airflow (therm). Hypopnea: ↓Airflow (therm) discernible with ↓SaO ₂ ≥ 4% or arousal	Apno-screen I [SaO ₂ ; body position; acti; HR]	Auto: ND Manual: Apnea: No airflow (therm). Hypopnea: ↓Airflow (therm) discernible with ↓SaO ₂ ≥ 4% or arousal	No	Tech & P (random)	Δt<4 wk	B
Pittman, 2004 USA	Referrals to a sleep lab	30 (29)	43	32 (7, 82) ^d	72	33.9	Apnea: No airflow (therm). Hypopnea: ↓Airflow (therm, Th/AB) ≥ 50% or "less reduction" with ↓SaO ₂ ≥ 3% or arousal	Watch Pat 100 [SaO ₂ ; HR; PAT; acti]	Auto: One of three: 1. ↓PAT amplitude with acceleration in pulse rate or ↑ wrist activity 2. ↓PAT amplitude with ↓SaO ₂ ≥3% (<4%) 3. ↓SaO ₂ ≥4%	ND	ND	mean Δt ^e = 1.7d	B
Bar, 2003 Israel	Referrals to a sleep center	14 (14)	ND	31 (4, 78) ^a	ND	ND	Apnea/Hypopnea: ↓Airflow (therm, Th/AB) ≥ 50% or "less reduction" with ↓SaO ₂ ≥ 3% or arousal	Watch Pat 100 [SaO ₂ ; HR; PAT; acti]	Auto: ND	No	P	ND	C

Author, Year Country	Partici-pants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/m ²)	Respiratory event definition for facility-based PSG	Name [Signals]	Scoring: Respiratory event definition	Portable	Att	Hook-up	Ti-ming	Quality
<i>Portable monitor assessed in hospital but not in a specialized sleep facility</i>														
Ayappa, 2004 USA	Referrals to sleep center & healthy people	66 (56)	ND	48 (4, 126)	58	38.2	Apnea: ↓Airflow (cannula, therm) ≥90% Hypopnea: ↓Airflow (Th/AB) ≥50% with ↓SaO ₂ ≥4% or arousal	Pro-Tech/ Compu-medics P2 [Airflow, SaO ₂ ; body position]	Manual: Apnea: ↓Airflow (cannula) ≥90% Hypopnea: ↓Airflow (Th/AB) ≥50% with ↓SaO ₂ ≥4% or arousal		No	Tech ^f	Δt ≤ 14d	A

Acti: actigraphy (wrist); Att: attended; Auto: automated scoring AHI: apnea-hypopnea index in events/hour of sleep; d: days; EEG: electroencephalogram; HR: heart rate; N_A/N_E: Number analyzed/enrolled; ND: not described; PAT: peripheral arterial tonometry; PSG: polysomnography; SaO₂: O₂ saturation; Tech: technologist; Th/AB belts: thoracoabdominal belts; therm: Thermistors used in airflow estimation; wk: weeks; y: year(s); Δt: time interval between the two studies. Respiratory events across all studies were of at least 10 seconds duration unless otherwise noted. Studies ordered by overall quality, and then by sample size after grouping per type IV monitor.

^a Data from digitized graph.

^b Evaluation of the duration and degree of desaturations and their trends using a complex algorithm.

^c Before (n=57) or after (n=134) facility-based PSG as scheduling permitted. No data on time interval.

^d Using the "Chicago criteria" for scoring.

^e Home study was first in 17 cases.

^f In the majority of patients (n=52).

Table 12: Description of studies comparing type IV monitors (one or two bioparameters) with facility-based polysomnography in the home setting.

Author, Year, Country	Participants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/m ²)	Respiratory event definition for facility-based PSG	Portable	Quality
<i>Portable monitors assessed at home</i>									
Pang, 2006 USA	Referrals to a sleep lab	39 (32)	52	32 (0, 111)	44	35.7	Apnea: No airflow (therm). Hypopnea: ↓Airflow (therm) discernible, with ↓Effort≥30% and ↓SaO ₂ ≥4% or arousal	Sleep Strip [Airflow] Auto: Algorithm assessing sleep parameters from airflow signal	No P 1d B
Baltzan, 2000 Canada	Referrals to a sleep lab	108 (74)	52	18 (ND)	74	28.4	Apnea: ↓ Airflow (therm)>90% Hypopnea: ↓Airflow >50% with ↓ SaO ₂ ≥4%	OxiFlow [SaO ₂ ; airflow] Auto: Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) >50% with ↓ SaO ₂ ≥4% ^a Manual: NA [Repetitive transient ↓SaO ₂ with rapid return to baseline with HR and airflow abnormalities]	No P Δt<2 wk C
Series, 1993 Canada	Referrals to sleep clinic	240 (240)	51	19 (ND)	90	31.7	Apnea: No airflow (therm) Hypopnea: ↓ SaO ₂ ≥4%	Oximeter ^b Manual: Transient ↓SaO ₂ (any degree) followed by a restoration to baseline	No P Δt<4 wk ^c B
Wiltshire, 2001 UK	Referrals	100 (84)	ND	16 (4, 111)	ND	ND	ND	Oximeter ^d Manual: [SaO ₂ dips]	No P Δt<3 d C
Ryan, 1995 UK	Referrals to sleep clinic	100 (69)	48	ND	83	29.6	Apnea: No airflow (therm) with ↓SaO ₂ ≥4% in next 30s Hypopnea: ↓Airflow (therm) ≥25% with paradoxical movement, ↓effort (Th) ≥25%, and ↓airflow (AB) ≥15%	Oximeter ^b Manual: NA ^e	No ND Δt= ND ^c C
Williams, 1991 USA	Referrals to sleep center	40 (ND)	55	29 (1-135)	ND	29.0	Apnea: No airflow (therm) Hypopnea: (not assessed)	Oximeter ^f Manual: ↓SaO ₂ ≥4% from baseline and absolute >92%	ND ND Δt<1 wk ^c C

Author, Year, Country	Parti-cipants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/m ²)	Respiratory event definition for facility-based PSG	Portable	Quality				
<i>Portable monitor assessed in an unclear setting</i>													
Bonsignore, 1990 Italy	OSAHS ^g and healthy controls	83 (83)	ND	ND	76	ND	ND	Oximeter	Manual: ↓SaO ₂ ≥ 4% until SaO ₂ returns to within 2% of baseline	No	ND	ND ^h	C

Att: attended; Auto: automated scoring AHI: apnea-hypopnea index in events/hour of sleep; d: days; N_A/N_E: Number analyzed/enrolled; ND: not described; OSAHS: Obstructive sleep apnea-hypopnea syndrome; PSG: polysomnography; SaO₂: O₂ saturation; Tech: technologist; Th/AB belts: thoracoabdominal belts; therm: Thermistors used in airflow estimation; wk: weeks; y: year(s); Δt: time interval between the two studies.

Respiratory events across all studies were of at least 10 seconds duration unless otherwise noted. Studies ordered by overall quality, and then by sample size after grouping per type IV monitor.

^a Alternative sets of criteria which were analyzed (apparently for at least 10 s duration): ↓25% in airflow with ↓4% SaO₂; ↓20% in airflow with ↓4% SaO₂, ↓50% in airflow with ↓2% SaO₂; ↓25% in airflow with ↓2% SaO₂.

^b Oximeter: Ohmeda Biox 3700.

^c Oximeter: Ohmeda Biox 3740.

^d Portable monitor was first.

^e British Thoracic Society criterion: OSAHS is diagnosed if awake baseline saturation is >90%, and there are at least 15 four ↓SaO₂>4% dips per hour in bed.

^f Oximeters: Ohmeda Biox 3700 and N100.

^g Case-control design. Cases had AHI >10 events/hour in facility-based PSG.

^h Facility-based PSG was first.

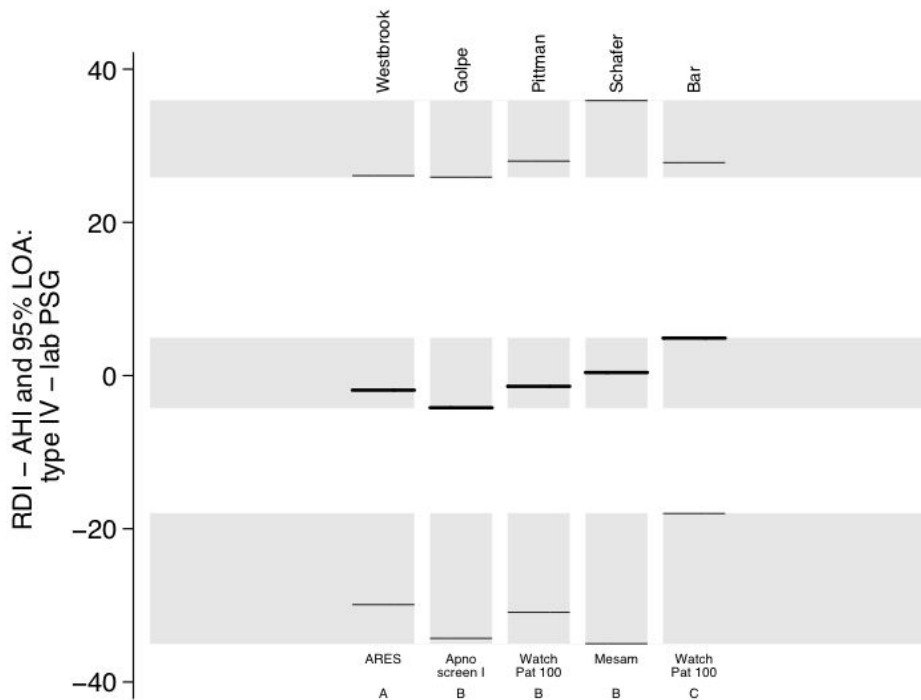
Findings of studies in the home setting—concordance

Type IV monitors that record 3 or more bioparameters (channels)

Figure 18 illustrates difference versus average analyses from five studies that reported this information. All five studies pertained to monitors that record at least three channels (see Definitions and Terminology and Section A3 for a relevant description and discussion). The average difference between the two measurements was between -4 to 5 events/hour. The graph shows the 95% limits of agreement in the measurements. The differences in the measurements were not dependent beyond chance on their average in three studies.^{59,105,123} (See the synopsis of Section B2 for an interpretation of this finding)

Not shown in **Figure 18** is the Ayappa 2004 study, that was performed in the hospital, but in a non-specialized sleep unit.⁷⁶ In that study, the portable monitors underestimated the AHI in facility-based PSG by -11 events/hour on average (95% limits of agreement: $-38, 17$). However, the differences were dependent on the average of the measurements, limiting the usefulness of this analysis.

Figure 18. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type IV monitors (with at least 3 channels) in specialized sleep units.



Schematic representation of the agreement between portable monitors and facility-based PSG as conveyed by difference versus average analyses. Each study is represented by three lines; these stand for the mean bias, and 95% limits of agreement from the difference versus average analyses. The upper, middle and lower grey areas group the upper 95% limits of agreement, the mean difference, and the lower 95% limits of agreement, respectively. The make of the monitor and the overall study quality are also depicted in the lower part of the graph.

Golpe 2002¹¹¹ found a good concordance between the Apnoscreen I monitor and facility-based PSG in classifying subjects into either with or without OSAHS (cutoff of 10 events/hour of sleep in facility-based PSG; $\kappa=0.78$ if the uncertain cases with

Apnoscreen I are classified as negative, or $\kappa=0.68$ if the uncertain cases are classified as positive for OSAHS). Similarly, Westbrook 2005 found $\kappa=0.77$ using the same threshold in facility-based PSG.¹²⁷

Type IV monitors that record one or two bioparameters (channels)

None of the eligible studies (**Table 12**) reported difference versus average analyses (or provided sufficiently clear graphs to allow digitizing).

Pang 2006 did not find evidence for agreement between individual measurements with the SleepStrip portable monitor and facility-based PSG ($\kappa=0.14$).

The study by Wiltshire 2001 primarily compared the number of dips in O₂ saturation per hour in bed between the portable oximeter (data storage every 12 seconds) and the oximeter used in the facility-based PSG (data storage every 2 seconds).⁴² They did not contrast patient classification with the oximeter versus all the information obtained by facility-based PSG, and thus do not contribute to any analyses.

Findings of studies in the home setting—predicting AHI suggestive of OSAHS

Ten studies^{59,81,105,106,111,118,122-124,127} assessed the ability of type IV monitors in the home setting to predict AHI in facility-based PSG that was considered suggestive of OSAHS. An AHI cutoffs of 15 events/hour of sleep was used as suggestive of OSAHS in six studies,^{59,81,105,118,122,123} 10 events/hour of sleep in five studies^{106,111,123,124,127} and 20 events/hour of sleep in two.^{123,124} Other cutoffs (5, 25, 30 or 40 events/hour) were also used in some of these studies.

Type IV monitors that record three or more bioparameters (channels)

Table 11 summarizes the characteristics studies that used monitors classified as type IV, although they record more than two channels (Refer to Terminology and Definitions and to Section A3 for more details). Five studies had relevant data.^{59,105,111,123,127} Using a cutoff of 15 events/hour of sleep in facility-based PSG, one out of three studies (Pittman 2004⁵⁹) had very high positive likelihood ratio and very low negative likelihood ratio (**Figure 19**). Using a cutoff of 10 events/hour, two out of three studies were near the region that implies low negative likelihood ratios.^{123,127} Schafer 1997¹²³ assessed the cutoff of 20 events/hour of sleep as well, finding a sensitivity of 68% and a specificity of 74%.

Type IV monitors that record one or two bioparameters (channels)

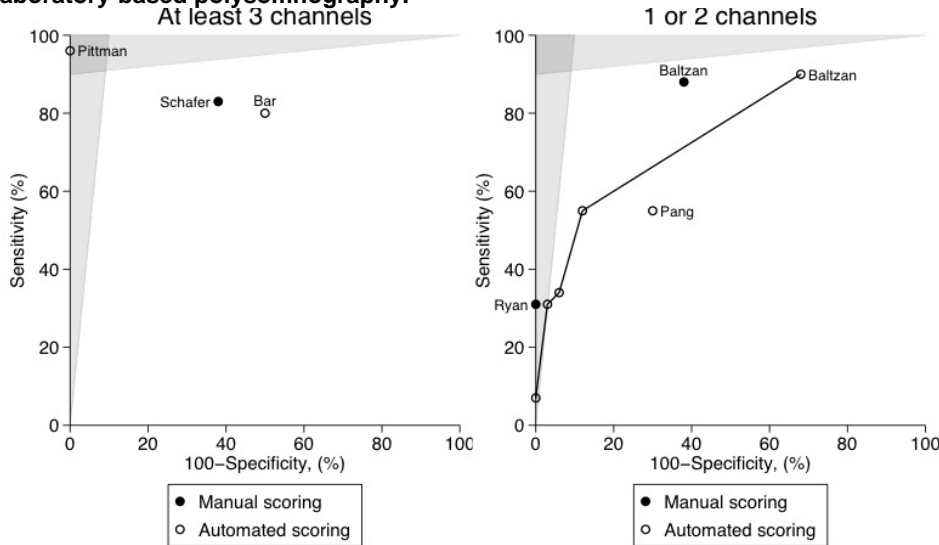
Table 12 summarizes the characteristics of the five studies that used oximeters or monitors recording only two channels (Refer to Terminology and Definitions and to Section A3 for more details).^{81,106,118,122,124}

Using a cutoff of 15 events/hour of sleep in facility-based PSG, none of the three studies with available data fall into the regions that defines both high positive and low negative likelihood ratios (**Figure 19**). Ryan 1995 and at least one cutoff in the Baltzan 2000 study had high positive likelihood ratios, but their sensitivity was well below 40%.

Two studies reported relevant data using a cutoff of 10 events/hour, namely Bonsignore 1990¹⁰⁶ and Series 1993¹²⁴ (**Figure 20**). For this cutoff, the first had a high positive likelihood ratio and the second had a low negative likelihood ratio.

Series 1993¹²⁴ assessed the cutoff of 20 events/hour of sleep as well, finding a sensitivity of 100% and a specificity of 61%.

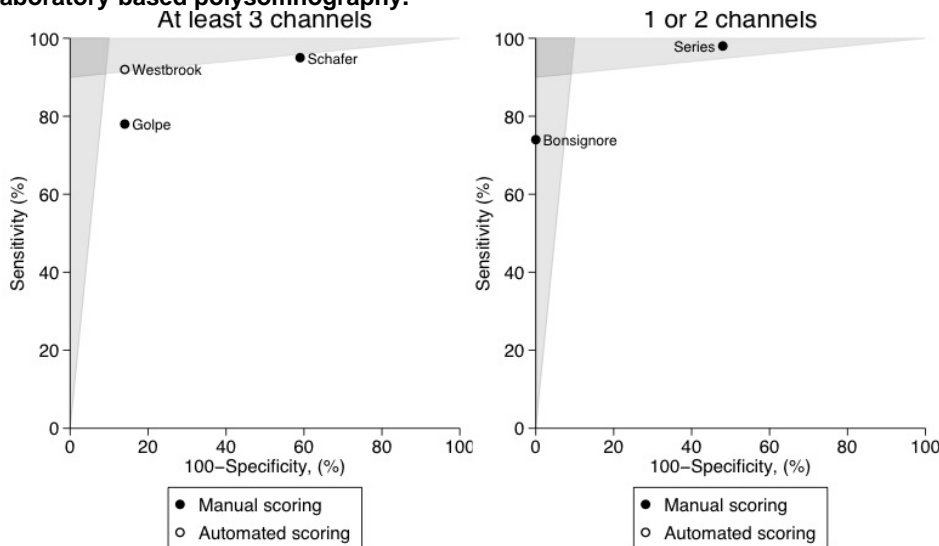
Figure 19. Diagnostic ability of type IV monitors in the home setting to identify AHI>15 events/hour in laboratory-based polysomnography.



Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded are have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both $LR+ > 10$ and $LR- < 0.1$. The line connects points representing different thresholds for the portable monitor from a single study.

Note that for Baltzan both the automated scoring and the manual scoring data are depicted.

Figure 20. Diagnostic ability of type IV monitors in the home setting to identify AHI > 10 events/hour in laboratory-based polysomnography.



Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded are have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both $LR+ > 10$ and $LR- < 0.1$.

More on type IV monitors in specialized sleep units

We identified 38 studies that were conducted in specialized sleep centers using monitors that were classified as type IV.^{26-30,42,58,59,80-84,103-105,107-110,112-117,119-121,125-127,129-134}

Table 13 summarizes the characteristics from 15 studies with monitors that record at least three channels. These would be “unclassified” under the original ASDA criteria (for a relevant discussion see “Terminology and Definitions” and Section A3 of this technology assessment). **Table 14** summarizes the characteristics of the remaining studies (assessing monitors that record only one or two bioparameters).

Across all studies, the average participant age was 50 years on median (interquartile range: 46, 52). The median number of analyzed subjects was 63 (interquartile range: 34, 140). Over 60% of the participants were males in the studies that reported gender distributions (**Table 13**). Participants were referral cases to sleep laboratories or specialized sleep centers, with the exception of the Gurubhagavatula 2004¹¹² study, which took sample from people with clinical suspicion of OSAHS from the general population.

Only seven papers clearly reported that the sleep monitor was attended by a technologist.^{30,58,81,110} In one study,¹⁰⁹ the oximeter was attended by nursing staff but in an inconsistent manner (and thus it was classified as unattended). It is likely that the technologist affixed the portable monitor’s probes in the majority of the studies. A technologist was clearly reported to had affixed the probes in eleven papers (**Table 13** and **Table 14**), and in Wiltshire 2001 the study participants did so.

Table 13. Description of studies comparing type IV monitors (recording three or more bioparameters) with facility-based polysomnography in specialized sleep units.

Author, Year, Country	Participants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/m ²)	Respiratory event definition for facility-based PSG	Name	Portable Scoring: Definition of respiratory events or other description	Att	Hook-up	Quality
Westbrook, 2005 USA	Referrals to sleep center	299 (284)	48	27 (1, 118)	62	ND	“Standard criteria”	ARES [SaO ₂ ; head position; snoring]	Auto: ↓SaO ₂ >2.2% or ↓SaO ₂ ≤2.2% with ↑SaO ₂ at 2.2% with arousal (based on activity) ^a	ND	ND	A
Pillar, 2003 Israel	Referrals to sleep center	68 (68)	46	34 (1, 118)	79	28.0	“Standard criteria”	Watch Pat 100 [SaO ₂ ; HR; PAT; acti]	Auto: NA [Attenuation of PAT signal amplitude, short movements]	Yes	ND	B
Ayas, 2003 US	Suspected OSAHS	30 (30)	47	23 (1-94)	63	31.0	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% or ↓airflow <50% with ↓SaO ₂ ≥3% or arousal	Watch Pat 100 [SaO ₂ ; HR; PAT; acti]	Auto: One of three: ↓PAT amplitude with acceleration in HR or ↑wrist activity ↓PAT amplitude with ↓SaO ₂ ≥3% (<4%) ↓SaO ₂ ≥4%	ND	ND	B
Pittman, 2004 USA	Referrals to sleep lab	30 (29)	43	32 (7, 82) ^b	72	30.9	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% or ↓airflow <50% with ↓SaO ₂ ≥3% or arousal	Watch Pat 100 [SaO ₂ ; HR; PAT; acti]	Auto: One of three: ↓PAT amplitude with acceleration in HR or ↑wrist activity ↓PAT amplitude with ↓SaO ₂ ≥3% (<4%) ↓SaO ₂ ≥4%	ND	ND	B
Bar, 2003 Israel	Referrals to sleep center; healthy	102 (99)	41	26 (2, 94) ^c	76	26.8	Apnea/hypopnea: ↓Airflow (therm) ≥50% or ↓airflow discernible with less reduction with arousal or ↓SaO ₂ ≥3%	Watch Pat 100 [SaO ₂ ; HR; PAT; acti]	Auto: [PAT signal amplitude, heart rate, and oxygen saturation; sleep/wake detected by wrist actigraphy]	No	ND	C
Penzel, 2004 Germany	Referrals to sleep lab	21 (17)	ND	15 (0, 84) ^c	ND	ND	Apnea: No airflow (therm) Hypopnea: ↓Effort (Th/AB) ≥50%	Watch Pat 100 [SaO ₂ ;	Auto: ↓PAT ≥40% with HR changes (no SaO ₂)	ND	Tech	C

Author, Year Country	Participants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/m ²)	Respiratory event definition for facility-based PSG	Name	Portable Scoring: Definition of respiratory events or other description	Att	Hook-up	Quality
Michaelson, 2006 USA	Referrals to sleep lab	59 (59)	40	15 (1, 80) ^{b,c}	83	26.6	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% or ↓airflow <50% with ↓SaO ₂ ≥3% or arousal	HR; PAT; acti] SNAP [SaO ₂ ; airflow; snoring; no other SNAP channel]	Auto + manual edit: Apnea: No sound Hypopnea: ↓≥75% sound amplitude with ↓SaO ₂ ≥4%	ND	ND	B
Esnaola, 1996 Spain	Referrals to sleep center	152 (150)	57	27 (ND)	89	29.8	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) discernible with ↓SaO ₂ ≥4% or arousal	Mesam IV [SaO ₂ ; HR; body position; snoring]	Auto: NA [ODI; snoring index; HR variation index] Manual: ↑HR>10%, ↓SaO ₂ ≥4%, and 3 snores separated by 10-120s	ND	ND	B
Koziej, 1994 Poland	Referrals to sleep lab	56 (56)	47	37 (0, 118)	91	ND	Apnea: No airflow (therm) Hypopnea: ↓Effort (Th/AB) ≥50%	Mesam IV [SaO ₂ ; HR; body position; snoring]	Auto: NA [ODI; snoring index; HR variation index] Manual: ND	ND	Tech	C
Stoohs, 1992 USA	Patients from sleep clinic	56 (56)	47	ND	82	27.0	“Standard criteria”	Mesam IV [SaO ₂ ; HR; body position; snoring]	Auto: ↓SaO ₂ ≥3%	ND	Tech	C
Rauscher, 1991 Austria	Unclear	53 (53)	50	19 (1, 88)	ND	ND	Apnea: No airflow (therm) Hypopnea: ↓Effort (Th/AB) ≥50% with ↓SaO ₂ ≥2% (if baseline absolute ≥94%) or ↓SaO ₂ ≥2% (if baseline absolute <94%)	Mesam [SaO ₂ ; EOG; snoring]	Auto: [Various indices: periods of constant HR between 11-60s; snoring periods between 11-60s; ↓SaO ₂ ; rapid ↑SaO ₂]	ND	ND	C
Guylay, 1987 Switzerland	Known OSAHS	14 (12)	ND	44 (3, 79)	100	27.9	Apnea: tidal volume <1/3 of resting for ≥15s Hypopnea: tidal volume between 1/3 and 2/3 of resting for ≥15s	Vitalog PMS-8 [SaO ₂ ; acti; induct pleth]	Manual: Apnea: tidal volume <1/3 of resting for ≥15s Hypopnea: tidal volume between 1/3 and 2/3 of	ND	ND	B

Author, Year, Country	Participants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/m ²)	Respiratory event definition for facility-based PSG	Name	Portable Scoring: Definition of respiratory events or other description	Att	Hook-up	Quality
Issa, 1993 Canada	Referrals to sleep center	129 (129)	48	ND	78	30.9	Apnea: No airflow (therm) Hypopnea: ↓Effort (Th/AB)≥50% with ↓SaO ₂ ≥3%	SnoreSat [SaO ₂ ; snoring]	resting for ≥15s Auto: ↓SaO ₂ >3% and amplitude of snoring ^d Manual: ↓SaO ₂ >3%	ND	Tech	B
van Surell, 1995 France	Referrals to sleep lab	50 (50)	52	22 (0, 74) ^c	98	27.0	Apnea: No airflow (therm) Hypopneas: ↓Airflow (therm) ≥50% with EEG arousal	CID 102 [SaO ₂ ; body position; tracheal sound]	Auto: Apnea: tracheal silence Hypopneas: short tracheal silence (for ≥7s & <10s) with cyclic ↓SaO ₂ ≥4% and ↑SaO ₂ within 50s	ND	Tech	B
Overland, 2005 Norway	Referrals to sleep center	53 (53)	ND	23 (1-119)	ND	ND	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% or ↓airflow <50% with ↓SaO ₂ ≥3% or arousal or ↑P _{esoph}	Reggie [SaO ₂ ; airflow; P _{esoph} ; acti]	Auto: [Default settings] Manual: Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% or ↓airflow <50% with ↓SaO ₂ ≥3% or ↑P _{esoph}	ND	ND	C

Acti: actigraphy (wrist); Att: attended; Auto: automated scoring; AHI: apnea-hypopnea index in events/hour of sleep; d: days; EEG: electroencephalogram; HR: heart rate; induct pleth: inductance plethysmography; NA: Not applicable; N_A/N_E: Number analyzed/enrolled; ND: not described; ODI: O₂ desaturation index; OSAHS: Obstructive sleep apnea-hypopnea syndrome; PAT: peripheral arterial tonometry; P_{esoph}: esophageal pressure; PSG: polysomnography; s: seconds; Tech: technologist; Th/AB belts: thoracoabdominal belts; therm: Thermistors used in airflow estimation; wk: weeks; y: year(s).

Respiratory events across all studies were of at least 10 seconds duration unless otherwise noted. Studies ordered by monitor used, quality and sample size.

^a Evaluation of the duration and the degree of desaturations and their trends using a complex algorithm.

^b Assessed with the Chicago criteria.

^c Data obtained from digitized graph.

^d Events: ↓SaO₂≥3% (compared to 15% of SaO₂ over a 5 minute sliding window) with a snore 5s before or 15s after. Snores defined as consecutive snores within 10 to 120 s; solitary snores do not qualify.

Table 14. Description of studies comparing type IV monitors (recording one or two bioparameters) with facility-based polysomnography in specialized sleep units.

Author, Year, Country	Participants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/m ²)	Respiratory event definition for facility-based PSG	Name	Portable Scoring: Definition of respiratory events or other description	Att	Hook-up	Quality
Mayer, 1998 France	Referrals to sleep lab	95 (95)	53	43 (1, 147) ^a	83	30.7	Hypopnea: ↓Airflow (therm) >50% with arousal or drop in SaO ₂ ≥4%	Autoset [SaO ₂ ; airflow]	Auto: Apnea/Hypopnea: ↓Airflow (cannula) ≥50%	No	Tech	B
Gugger, 1997 Switzerland	Referrals	67 (67)	51	26 (0, 96) ^a	87	31.0	Apnea: ↓No airflow (therm) Hypopnea: ↓Airflow (therm) >50%	Autoset [SaO ₂ ; airflow]	Auto: Apnea: ↓Airflow (cannula) ≥75% Apnea/Hypopnea: ↓Airflow (cannula) ≥50%	Yes	Tech	B
Bagnato, 2000 Brazil	Suspected OSAHS	63 (56)	45	38 (ND)	80	31.3	Apnea: ↓Airflow (therm) ≥80% Hypopnea: ↓Airflow (therm) discernible with ↓SaO ₂ ≥4% or arousal	Autoset [SaO ₂ ; airflow]	Auto: Apnea/Hypopnea: ↓Airflow (therm) ≥50%	Yes	Tech	B
Kiely, 1996 Ireland	ND	41 (36)	44	19 (ND)	75	28.0	Apnea: ↓Airflow (therm) >80% Hypopnea: ↓Airflow (therm) >50%	Autoset [SaO ₂ ; airflow]	Auto: Apnea: ↓Airflow (cannula) ≥75% Apnea/Hypopnea: ↓Airflow (cannula) ≥50%	Yes	Tech	B
Fleury, 1996 France	ND	44 (38)	52	ND	77	28.5	Apnea: No airflow (therm)	Autoset [SaO ₂ ; airflow]	Auto: ND (defaults?) ^b	ND	ND	C
Bradley, 1995 England	ND	31 (31)	46	20 (2, 81) ^a	84	30.0	“Standard criteria”	Autoset [SaO ₂ ; airflow]	Auto: Apnea/Hypopnea: ↓Airflow (cannula) >50%	ND	ND	C
Gugger, 1995 Switzerland	ND	27 (27?)	51	ND	85	29.0	“Standard criteria”	Autoset [SaO ₂ ; airflow]	Auto: ^c Apnea: No airflow (cannula)	ND	ND	C
Rees, 1998 UK	Referrals to sleep lab	20 (20)	48	39 (8-114)	100	31.0	Apnea: No airflow (therm) Hypopnea: ↓Effort (Th/AB) ≥50%	Autoset [SaO ₂ ; airflow]	Auto: Apnea: ↓Airflow (cannula) ≥75% Apnea/Hypopnea: ↓Airflow (cannula) ≥50%	ND	ND	C
Baltzan, 2000	Referrals to sleep	108 (86)	52	18 (ND)	74	28.4	Apnea: ↓Airflow (therm) ≥90%	OxiFlow [SaO ₂ ;	Auto: Apnea: No airflow (therm)	Yes	Tech	C

Author, Year, Country	Participants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/m ²)	Respiratory event definition for facility-based PSG	Name	Portable Scoring: Definition of respiratory events or other description	Att	Hook-up	Quality
Canada	lab						Hypopnea: ↓Airflow ≥50% with ↓SaO ₂ ≥4%	airflow]	Hypopnea: ↓Airflow (therm) >50% with ↓SaO ₂ ≥4% ^d Manual: NA [Repetitive transient ↓SaO ₂ with rapid return to baseline with HR and airflow abnormalities]			
Shochat, 2002, Israel, Belgium, Germany	Suspected OSAHS	402 (288)	ND	ND	ND	ND	Apnea: No airflow (therm) Hypopnea: ↓Effort (Th/AB) ≥50% with ↓SaO ₂ ≥4%	Sleep-Strip [Airflow]	Auto: ND ^e	ND	ND	C
Gurubhagavatula, 2004 USA	Suspected OSAHS ^f	406 (406)	44	ND	94	28.4	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥3% or arousal	Oximeter	Manual: ↓SaO ₂ ≥3%	ND	ND	B
Levy, 1996 France	Referrals to sleep lab	301 (301)	56	30 (ND)	ND	32.0	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50%	Oximeter ^g	Auto: NA [δ index]	ND	ND	B
Zamarron, 2003 Spain	Referrals to a sleep clinic	314 (300)	57	40 (ND)	78	29.5	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) [discernible?] with ↓SaO ₂ ≥4%	Oximeter ^h	Auto: NA [Peak in the periodogram between 30 and 70 s]	ND	ND	B
Chiner, 1999 Spain	Referrals to sleep center	275 (275)	51	42 (15-101)	89	30.0	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% or ↓effort (Th/AB) ≥50% or ↓airflow ≥50%; both with ↓SaO ₂ ≥4% or arousal	Oximeter ⁱ	Manual: ↓SaO ₂ ≥4% of previous minute average	ND	ND	B
Zamarron, 1999 Spain	Referrals to sleep clinic	240 (233)	57	40.1 (ND)	80	30.4	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% with ↓SaO ₂ ≥4%	Oximeter ^h	Auto: NA [Peak in the periodogram between 30 and 70 s]	ND	ND	B
Adachi, 2003 Japan	Referral to sleep unit	33 (31)	49	32	90	ND	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) ≥50% or ↓airflow <50%	Oximeter ^j	Auto: NA [Algorithm estimating a pulse rate rise index]	ND	ND	C

Author, Year Country	Participants	N _E (N _A)	Mean age (y)	Mean AHI (Range)	Male (%)	Mean BMI (kg/m ²)	Respiratory event definition for facility-based PSG	Name	Portable Scoring: Definition of respiratory events or other description	Att	Hook-up	Quality
Vazquez, 2000 Canada	Referrals to sleep center	245 (241)	45	26 (0-132)	78	30.8	with ↓SaO ₂ ≥3% or arousal Apnea: No airflow (therm) Hypopnea: (A) ↓SaO ₂ ≥4% (B) ↓SaO ₂ ≥4% or arousal	Oximeter ^k	Auto: ↑SaO ₂ preceded by ≥3 consecutive ↓SaO ₂ (one of them >4%)	Yes	Tech	C
Douglas, 1992 UK	Referrals to sleep lab	220 (200)	50	ND (0, 95) ^a	82	ND	ND	Oximeter ^g	Manual: ↓SaO ₂ ≥4% [Also ↓SaO ₂ ≥2% or 3%]	Yes	ND	C
Alvarez, 2006 Spain	Referrals	187 (187)	58	40 (ND)	79	29.5	Apnea: No airflow (therm) Hypopnea: ↓Airflow (therm) discernible with ↓SaO ₂ ≥4%	Oximeter ^h	Auto: NA [δ index; CTM; Lempel Ziv complexity] ^l	ND	ND	C
Rauscher, 1993 Austria	Referrals to sleep lab	116 (116)	ND	ND	82	ND	Apnea: No airflow (therm) Hypopnea: ↓Effort (Th/AB) ≥50% with ↓SaO ₂ ≥2% (if baseline absolute ≥94%) or ↓SaO ₂ ≥2% (if baseline absolute <94%)	Oximeter ^l	Manual: NA [Patterns of cyclic oscillations in SaO ₂ or in HR for >30 min]	ND	ND	C
Bradley, 1995 England	ND	31 (31)	46	20 (2, 81) ^a	84	30.0	"Standard criteria"	Oximeter ^g	Manual: NA [Visual categorization as definitely indicative of OSAHS or not]	ND	ND	C
Cooper, 1991 UK ^m	Referrals to sleep center	45 (41)	ND	ND	63	ND	ND	Oximeter ^g	Manual: NA [Repetitive ↓SaO ₂ ≥5%]	No ⁿ	ND	C
Gugger, 1995 Switzerland	ND	27 (27?)	51	ND	85	29.0	"Standard criteria"	Oximeter	Auto: Apnea: No airflow (cannula)	ND	ND	C
Pepin, 1991 USA	ND	26 (26)	59	51 (ND)	ND	ND	Apnea: No airflow (therm) Hypopnea: tidal volume <1/3 of previous breaths	Oximeter ^g	Auto: NA [δ index]	ND	ND	C
Wiltshire, 2001 UK	Referrals	100 (16)	ND	16 (4, 111)	ND	ND	ND	Oximeter ^g	Manual: [SaO ₂ dips]	No	P	C

Att: attended; Auto: automated scoring; AHI: apnea-hypopnea index in events/hour of sleep; CTM: Central tendency measure; d: days; N_A/N_E : Number analyzed/enrolled; ND: not described; ODI: O₂ desaturation index; OSAHS: Obstructive sleep apnea-hypopnea syndrome; PSG: polysomnography; s: seconds; Tech: technologist; Th/AB belts: thoracoabdominal belts; therm: Thermistors used in airflow estimation; wk: weeks; y: year(s)
Respiratory events across all studies were of at least 10 seconds duration unless otherwise noted. Studies ordered by monitor used, quality and sample size.

^a Data obtained from digitized graph.

^b In the last 15 participants the Autoset measured combined apneas and hypopneas, not apneas only.

^c Instead of version 3.03 of the Autoset software that was used in the other studies, a prototype version of the software was used in this study.

^d Alternative sets of criteria which were analyzed (apparently for at least 10s duration): ↓25% in airflow with ↓4% SaO₂; ↓20% in airflow with ↓4% SaO₂, ↓50% in airflow with ↓2% SaO₂; ↓25% in airflow with ↓2% SaO₂.

^e Assessed with the Chicago criteria.

^f Using general population sampling and selection by interview.

^g Oximeter: Ohmeda Biox 3700 or 3740 or IIA.

^h Oximeter: Critical care 504.

ⁱ Oximeter: Nellcor N-200.

^j Oximeter: Minolta Pulsox M24 or Pulsox 7.

^k Oximeter: Healthdyne 202-11.

^l CTM, (central tendency measure) counts how cases in the second order difference plot lie outside a circle with radius ρ ; however, here ρ is derived from the same set of patients (CMT is not independent validation, results may be largely upwardly biased). The Lempel-Ziv complexity measures the rate at which new patterns arrive when a signal is evaluated (does not suffer from the CTM derivation/validation problem).

^m Oximetry performed during the first night, and facility-based PSG during the second night.

ⁿ Nursing staff was overseeing the oximeter (portable), but not constantly.

Findings of studies in the specialized sleep unit setting—concordance

Type IV monitors that record at least three bioparameters (channels)

Figure 21, panel A, illustrates difference versus average analyses from relevant studies that received grade “A” or “B” for their overall methodological quality. The average difference between the two measurements ranged from –9 to 9 events/hour. The 95% limits of agreement were very wide. There was no statistically significant dependency of the difference of the measurements on their mean in only in one study.⁵⁹ In the other studies this information was not extractable^{28,126,127} or such a dependency existed.^{80,116} Therefore, the interpretation of difference versus average analyses should be done with caution (See the synopsis of Section B2 for an interpretation of such analyses).

Figure 22, panel A, illustrates the corresponding analyses for the studies that received grade “C” for their overall methodological quality.^{105,117,119,125,134} Again the two measurements may differ substantially for given individuals, as evident from the mean differences and the respective 95% limits of agreement in the graphs.

Westbrook 2005¹²⁷ found high agreement between the ARES monitor and facility-based PSG in classifying into those with or without OSAHS ($\kappa=0.85$; 10 events/hour of sleep were suggestive of OSAHS in facility-based PSG). Esnaola 1996²⁸ found high and modestly high intraclass correlation coefficients for manual and automated scoring (ICC=0.72 and ICC=0.47, respectively).

Other studies assessed the concordance between other arousal-related indices,³⁰ or described concordance as percentage of people where the difference between the two measurements was less than 30% of their average.¹²⁰ Overall, their assessment offered no additional insights.

Type IV monitors that record one or two bioparameters (channels)

Four studies using the Autoset monitor (records SaO₂ and airflow in its “diagnostic” mode) were of moderate methodological quality (**Figure 21**, panel B).¹²⁹⁻¹³² Although the same monitor was used, the mean bias ranges from –9.6 to 4.2 events/hour.

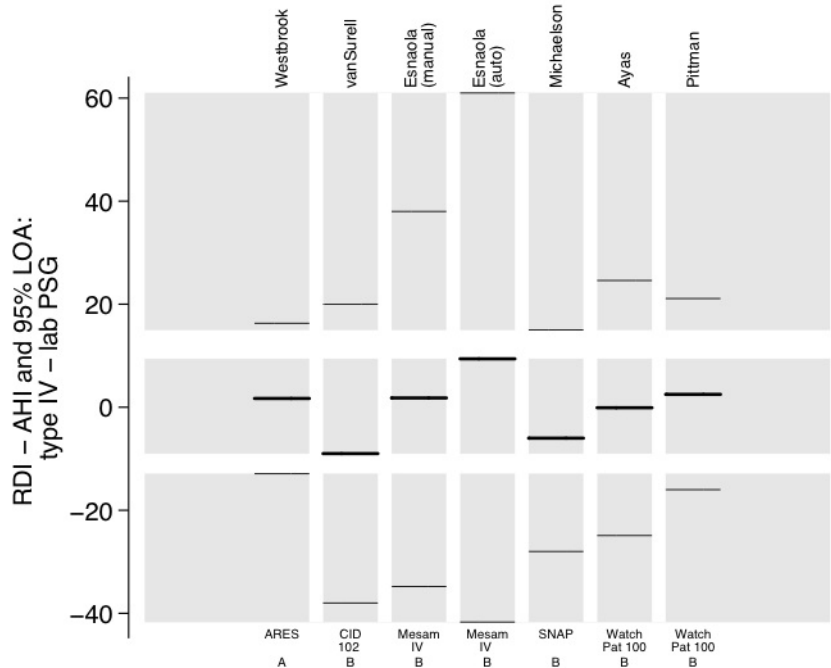
Five studies that were graded “C” for their overall methodological quality provided data for difference versus average analyses (**Figure 22**, panel B). In Rees 1998 the average difference was large (-17 events/hour), indicating that the Autoset monitor underestimated greatly the AHI from facility-based PSG.¹³³ The remaining three studies^{58,81,107} had mean bias between –0.1 and 5.4 events/hour, compared to facility-based PSG measurements.

Three studies compared measurements of the apnea index (AI).^{26,82,130} The results of the difference versus average analyses were similar. Portable monitors over estimated AI on average by 2.5 apneas/hour in Gugger 1997,¹³⁰ 2.6 apneas/hour in Fleury 1996,²⁶ and 8.9 apneas/hour in Gugger 1995.⁸²

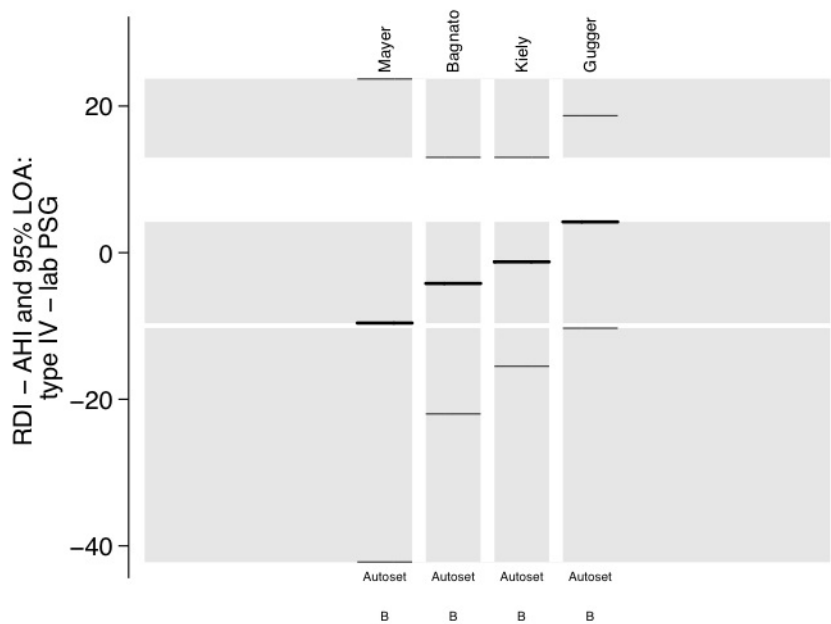
Bagnato 2000¹²⁹ found a very high intraclass correlation coefficient (ICC=0.94) between the AHI obtained from facility-based PSG and the RDI obtained from Autoset. Adachi 2003 also assessed the concordance in breath-related arousals (not reported in detail here because it did not offer additional insights – see Evidence Tables).¹⁰³

Figure 21. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type IV monitors (with at least 3 versus 1 or 2 channels) in specialized sleep units. Studies with good or moderate overall methodological quality.

A: At least 3 channels



B: 1 or 2 channels

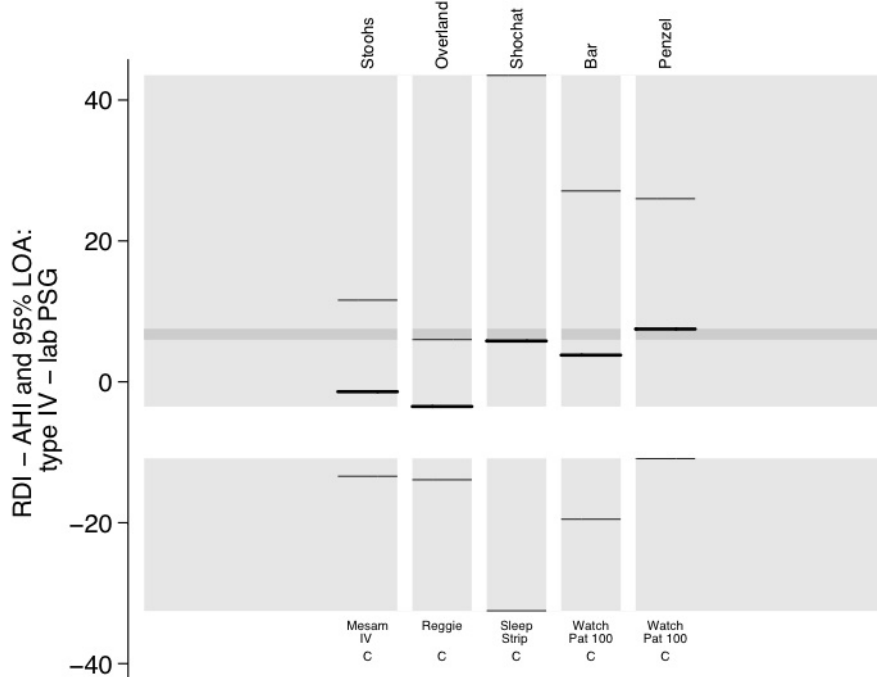


Type IV monitors that record at least three channels are depicted in panel A (upper graph); type IV monitors that record one or two channels are depicted in panel B (lower graph).

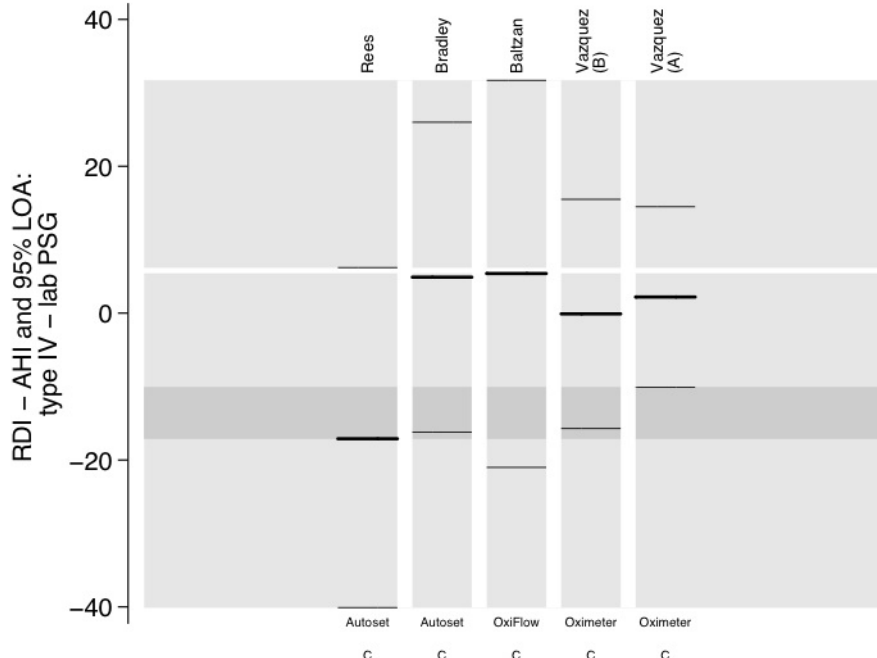
Schematic representation of the agreement between portable monitors and facility-based PSG as conveyed by difference versus average analyses. Each study is represented by three lines; these stand for the mean bias, and 95% limits of agreement from the difference versus average analyses. The upper, middle and lower grey areas group the upper 95% limits of agreement, the mean difference, and the lower 95% limits of agreement, respectively. The make of the monitor and the overall study quality are also depicted in the lower part of the graph. Only studies with good or moderate overall methodological quality are illustrated.

Figure 22. Schematic representation of the mean bias and limits of agreement between facility-based PSG and type IV monitors (with at least 3 versus 1 or 2 channels) in specialized sleep units. Studies with poor overall methodological quality.

A: At least 3 channels



B: 1 or 2 channels



Type IV monitors that record at least three channels are depicted in panel A (upper graph); type IV monitors that record one or two channels are depicted in panel B (lower graph).

Schematic representation of the agreement between portable monitors and facility-based PSG as conveyed by difference versus average analyses. Each study is represented by three lines; these stand for the mean bias, and 95% limits of agreement from the difference versus average analyses. The upper, middle and lower grey areas group the upper 95% limits of agreement, the mean difference, and the lower 95% limits of agreement, respectively. Note that the upper and middle grey areas overlap slightly. The make of the monitor and the overall study quality are also depicted in the lower part of the graph. Only studies with poor overall methodological quality are illustrated.

Vazquez used two definitions for hypopneas (A) and (B) in facility-based PSG (Table 14).

Findings of studies in the specialized sleep unit setting—predicting AHI suggestive of OSAHS

Thirty studies^{27-29,58,59,80-84,104,107-110,112,114-116,120,125,126,129-134} assessed the ability of type IV monitors in the lab setting to predict AHI in facility-based PSG that was suggestive of OSAHS (using a cutoff of 15 events/hour of sleep,^{28,58,59,80,81,107-110,114-116,126,129,131-133} 10 events/hour of sleep^{28,29,58,80,83,84,110,114,121,125,129,131,134} and 20 events/hour of sleep^{28,58,80,82,110,114,129-132,134}). Other thresholds were also assessed (5, 7, 25 and 30 events per hour of sleep). The study by Vazquez 2000⁵⁸ excluded participants who had measurements of less than 5 events per hour above or below the classification cutoff, and thus yielded misleadingly high sensitivity and specificity. It is not presented in the following analyses. Another study by Alvarez 2006¹⁰⁴ did not state the cutoff that was used in facility-based PSG and thus was not included in the pertinent graphs.

Type IV monitors that record three or more bioparameters (channels)

Table 13 summarizes the characteristics of fifteen studies that used monitors classified as type IV, although they record more than two channels (Refer to Terminology and Definitions and to Section A3 for more details).

In some studies the authors clearly report selecting cutoffs in the type IV monitor measurements so that the sensitivity or specificity would be high. This is likely to have happened in many studies, judging from the location of the points in the sensitivity/1-specificity plots. Overall, studies were on or very close to the regions that defined high positive likelihood ratios or low negative likelihood ratios to identify people with more than 15 events per hour in facility-based PSG.

Two studies provided sensitivity and specificity pairs using manual scoring and a cutoff of 15 events/hour of sleep in facility-based PSG.^{28,116} For both studies, the reported sensitivity and specificity pairs corresponded to a high positive likelihood ratio (**Figure 23**). Five studies used automated scoring for the same cutoff (15 events/hour) in facility-based PSG (**Figure 24**).^{28,59,80,114,126} With the exception of van Surell 1995,¹²⁶ all studies had high positive or low negative likelihood ratios (none had both). Four out of six studies with either manual or automated scoring reported sensitivity and specificity pairs that correspond to high positive or low negative likelihood ratios (no study had both high LR+ and low LR-).

Similar observations were made when a cutoff of 10 events/hour (**Figure 25** and **Figure 26**) and a cutoff of 20 events/hour were used in facility-based PSG (**Figure 27**). Using the cutoff of 10 events/hour, two out of five studies with either manual or automated scoring reported sensitivity and specificity pairs that correspond to high positive or low negative likelihood ratios (one study had both high LR+ and low LR-).

Type IV monitors that record one or two bioparameters (channels)

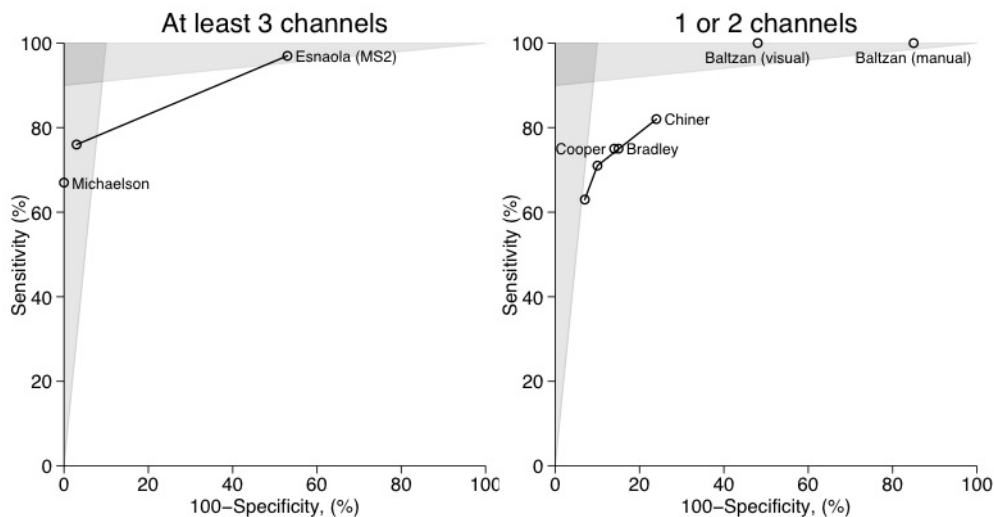
Table 14 summarizes the characteristics of the five studies that used oximeters or monitors recording only two channels (Refer to Terminology and Definitions and to Section A3 for more details).

Four studies provided sensitivity and specificity pairs using manual scoring and a cutoff of 15 events/hour of sleep in facility-based PSG.^{81,107-109} (Figure 23). Figure 24 depicts the three studies that used automated scoring.^{81,110,115} Using the cutoff of 15 events/hour, four out of six studies with either manual or automated scoring reported at least one sensitivity and specificity pair that corresponds to high positive or low negative likelihood ratios (no study had both high LR+ and low LR-).

Rausher 1991¹²⁰ had extractable sensitivity and specificity for using manual scoring and a cutoff of 10 events/hour of sleep in facility-based PSG (Figure 25). Three additional studies reported results for automated scoring with the same cutoff.^{83,84,110} Two out of four studies with either manual or automated scoring were in regions that correspond to high positive or low negative likelihood ratios (none had both high LR+ and low LR-).

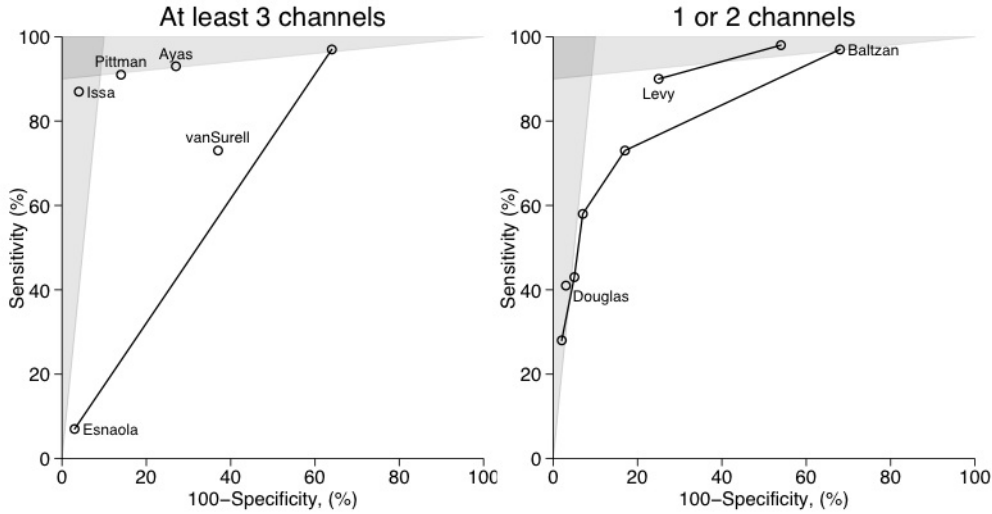
Figure 26 shows the two studies that used a cutoff of 20 events/hour in facility-based PSG.

Figure 23. Diagnostic ability of type IV monitors (with at least 3 versus 1 or 2 channels) in specialized sleep centers to identify AHI>15 events/hour in laboratory-based polysomnography. Studies using manual scoring for the type IV monitor.



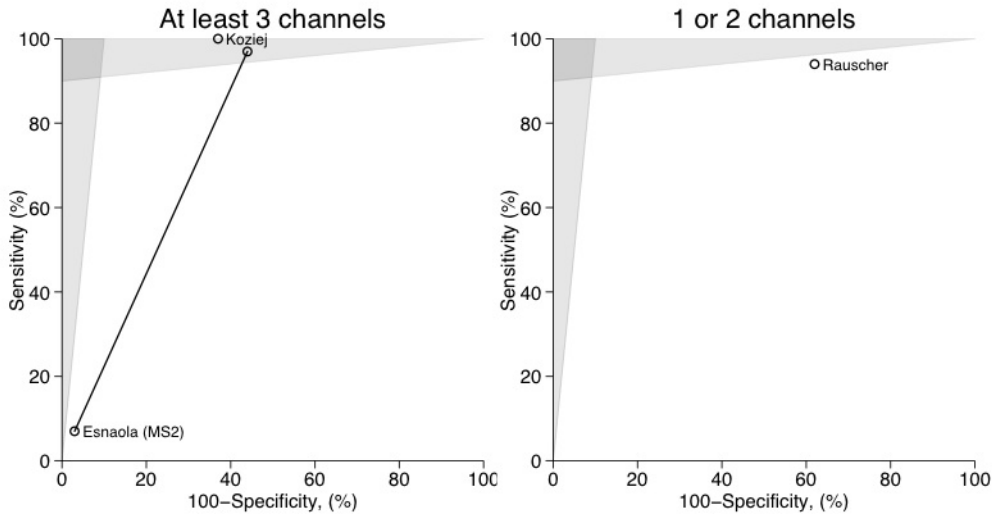
Sensitivity/specificity pairs from the same study (obtained with different cutoffs for the type IV monitor) are connected with lines. These lines do not represent the ROC curves from the pertinent studies. Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded area have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both LR+>10 and LR-<0.1. Baltzan (visual) used a binary classification of type IV recordings as suggestive of OSAHS or not. Baltzan (manual) used a manual counting of the number of respiratory events with the type IV monitor (Table 14).

Figure 24. Diagnostic ability of type IV monitors (with at least 3 versus 1 or 2 channels) in specialized sleep centers to identify AHI>15 events/hour in laboratory-based polysomnography. Studies using automated scoring for the type IV monitor.



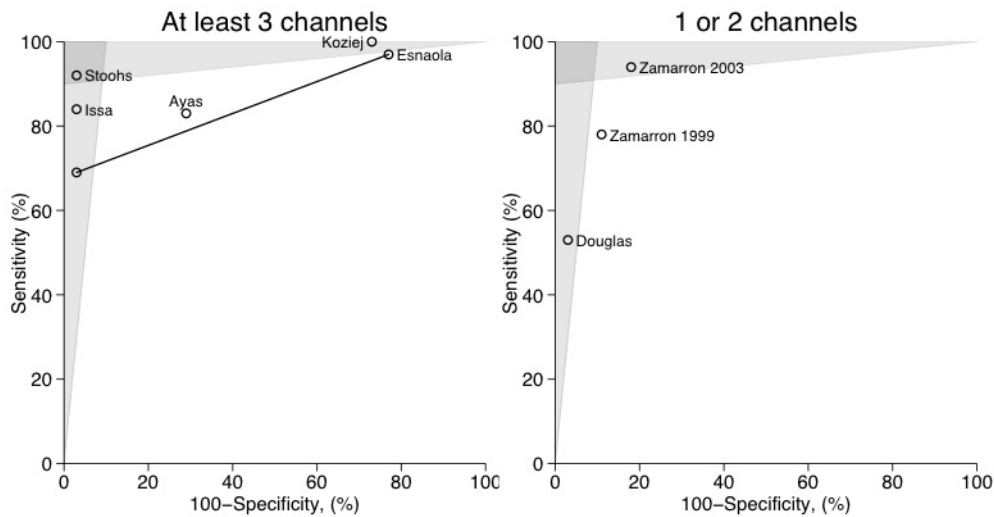
Sensitivity/specificity pairs from the same study (obtained with different cutoffs for the type IV monitor) are connected with lines. These lines do not represent the ROC curves from the pertinent studies. Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded are have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both $LR > 10$ and $LR < 0.1$.

Figure 25. Diagnostic ability of type IV monitors (with at least 3 versus 1 or 2 channels) in specialized sleep centers to identify AHI>10 events/hour in laboratory-based polysomnography. Studies using manual scoring for the type IV monitor.



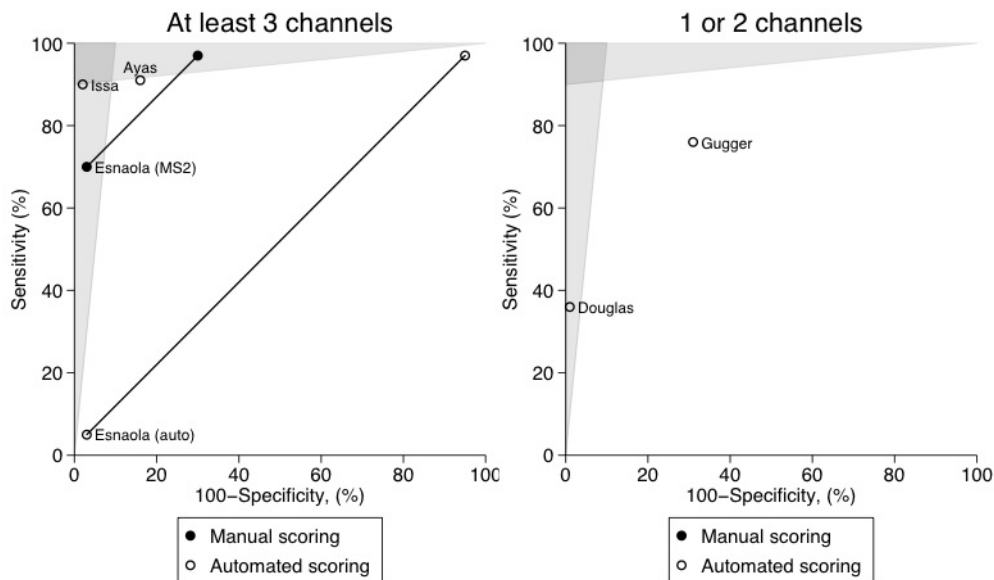
Sensitivity/specificity pairs from the same study (obtained with different cutoffs for the type IV monitor) are connected with lines. These lines do not represent the ROC curves from the pertinent studies. Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded are have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both $LR > 10$ and $LR < 0.1$. Esnaola 1996 used two manual scoring methods. The one that needed two and not three valid signals to characterize an event is shown.

Figure 26. Diagnostic ability of type IV monitors (with at least 3 versus 1 or 2 channels) in specialized sleep centers to identify AHI>10 events/hour in laboratory-based polysomnography. Studies using automated scoring for the type IV monitor.



Sensitivity/specificity pairs from the same study (obtained with different cutoffs for the type IV monitor) are connected with lines. These lines do not represent the ROC curves from the pertinent studies. Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded are have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both $LR+ > 10$ and $LR- < 0.1$.

Figure 27. Diagnostic ability of type IV monitors (with at least 3 versus 1 or 2 channels) in specialized sleep centers to identify AHI>20 events/hour in laboratory-based polysomnography.



Sensitivity/specificity pairs from the same study (obtained with different cutoffs for the type IV monitor) are connected with lines. These lines do not represent the ROC curves from the pertinent studies. Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded are have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both $LR+ > 10$ and $LR- < 0.1$.

Esnaola 1996 used two manual scoring methods. The one that needed two and not three valid signals to characterize an event is shown.

Synopsis for section B2

Difference versus average analyses suggest that substantial differences in the AHI may be encountered between type II monitors and facility-based PSG, especially when the studies are not performed simultaneously, but at different nights. Even larger differences compared with facility-based PSG cannot be excluded for type III monitors; and more so for type IV monitors. This was true both for manual and automated scoring.

Based on limited data, type II monitors may identify AHI suggestive of OSAHS with high positive likelihood ratios and low negative likelihood ratios. Type III monitors may have the ability to predict AHI suggestive of OSAHS with high positive likelihood ratios and low negative likelihood ratios for various AHI cutoffs in laboratory-based PSG. Studies of type IV monitors that record at least three bioparameters showed high positive likelihood ratios and low negative likelihood ratios. Studies of type IV monitors that record one or two bioparameters also had high positive likelihood ratios and low negative likelihood ratios, at least for selected sensitivity and specificity pairs from ROC curve analyses.

Overall, the ability of portable monitors to predict AHI with facility-based PSG appears to be worse in studies conducted in the home setting compared to studies in the specialized sleep laboratory. Between-night variability is a plausible explanation: in the sleep-lab setting measurements are simultaneous, whereas in other settings (home, hospital ward, outside the sleep lab) measurements are performed in different nights.

Interpretation

Apparent discrepancies between difference versus average analyses and sensitivity and specificity analyses. The observation that the mean bias and 95% limits of agreement cannot exclude substantial differences in the AHI between portable monitors and facility-based PSG is not incompatible with the calculated high positive likelihood ratios or low negative likelihood ratios.

In almost all studies where difference versus average analyses were undertaken, the differences between portable monitors and facility-based PSG were more pronounced for large AHI or RDI levels, and smaller for lower AHI or RDI levels. For this reason, the mean bias and 95% limits of agreement do not describe the whole range of AHI or RDI measurements. Therefore, the portable monitors may still be able to predict which people would have AHI in PSG more than 15 events/hour (a relatively small value in the whole range of AHI values that often range to over 100 events/hour). Furthermore, as discussed in Section A1 and Section B1, the exact AHI values may not convey additional useful information for people who have severe OSAHS (large AHI values).

Difference versus average analyses stress that the individual RDI or AHI values from portable monitors and facility-based PSG are not interchangeable (especially for large AHI or RDI levels).

The meaning of sensitivity, specificity, and related analyses. The analyzed studies used measurements from facility-based PSG that were *suggestive of OSAHS* as a reference standard. We note that the actual *diagnosis of OSAHS* is not set solely on the basis of high AHI values (although several studies made such assumptions). Section A1 provides a discussion of why AHI, or other related indices from facility-based PSG, are not an error-free reference standard for the diagnosis of OSAHS.

Therefore, the sensitivity, specificity, and likelihood ratios of portable sleep monitors refer only to their ability to predict AHI above a given cutoff in facility-based PSG. Whether a strong association with a true diagnosis of OSAHS exists or not, is dependent on several factors:

1. The actual AHI cutoff that was used. It may be argued that a cutoff of 15 events/hour in facility-based PSG would be more suggestive of OSAHS compared to a cutoff of 10 events/hour.
2. The presence of conditions that affect sleep quality and may not be equally identified by facility-based PSG and some portable monitors.
 - a. People with cardiac insufficiency or atrial flutter may exhibit Cheyne-Stokes breathing patterns (periodic apneas with a central component). The same may be true for people with respiratory disorders (e.g., chronic obstructive pulmonary disease, obesity hypoventilation syndrome) or sleep disorders (e.g. narcolepsy). This may pose difficulties in the diagnosis of OSAHS, especially with portable monitors that do not record effort channels.
 - b. Periodic Limb Movements in Sleep (PLMS) are observed in the majority of patients with the Restless Leg Syndrome (RLS).¹³⁵ These are involuntary clonic-type movements of the lower extremities while sleeping that may result in multiple arousals and disrupt sleep.¹³⁶ Overnight oximetry (type IV monitor with one bioparameter) would not easily differentiate PLMS from OSAHS. This may not be the case with other monitors that record airflow information.
3. The prior probability of OSAHS. People who have been referred for sleep studies because of suggestive symptoms are more likely to have OSAHS. Among them, an AHI suggestive of OSAHS is strongly associated with a true diagnosis of OSAHS. People with high prior probability of OSAHS would be:
 - a. Referrals from physicians familiar with OSAHS and its differential diagnosis
 - b. People with symptoms or suggestive clinical profile (e.g., middle aged males with high BMI)

Applicability to the Medicare population. Three caveats may be made:

First, the prior probability of OSAHS among Medicare beneficiaries may be lower than the prior probability of OSHS among subjects analyzed in the included studies:

1. The average age of the participants in the analyzed studies was recorded. The median value was 50 to 52 years. Moreover, the majority of the subjects were males and they had BMI above 25 kg/m². Finally, in most studies subjects were referred for PSG by sleep or respiratory physicians. On the other hand, Medicare beneficiaries are older (≥ 65 years), are not predominantly male, and may often have comorbidities that affect sleep quality.
2. There is evidence that, in the elderly, obesity and daytime somnolence may not be as strongly associated with OSAHS as in middle aged people.¹⁵ In other words, a high AHI or RDI value would not be as strongly associated with an *actual diagnosis of OSAHS*.

Second, facility-based PSG can differentiate OSAHS from other conditions that cause sleep disturbances (such as PLMS/RLS, or conditions associated with mixed or central apneas or hypopneas). The following caveats may be made:

1. Conditions that are associated with sleep disturbances with a central component (such as e.g., cardiac insufficiency) are prevalent among Medicare beneficiaries. This may be true for respiratory conditions (chronic obstructive pulmonary disease, obesity hypoventilation syndrome and other hypoventilatory syndromes) as well as other sleep disorders like narcolepsy. This caveat probably pertains to type III and type IV monitors.
2. Oximeters (and some monitors that do not assess airflow or effort) may not be able to differentiate PLMS/RLS from OSAHS. The prevalence of RLS is increased among older adults (up to 8% in people older than 60 years, compared to approximately 5% in people who are in their forties).¹³⁷ Moreover, PLMS/RLS is more common in women,¹³⁷ who were the minority in the studied populations. This caveat pertains to some type IV monitors that assess one or two bioparameters only.

Finally, in almost all eligible studies the investigators were associated with sleep clinics and/or sleep laboratories. Study investigators are probably very familiar with the diagnosis and treatment of sleep disorders (including ones other than OSAHS) even when this is not clearly stated in the primary papers). One cannot necessarily extrapolate the findings in these studies to circumstances where health care providers with less training and experience might use the devices.

B2a. For studies in the home setting, do technologist support and patient education affect the comparison of portable monitors and with facility-based polysomnography for the diagnosis of OSAHS?

This question is discussed separately per type of monitor.

Type II monitors

Three studies of type II monitors were performed in the home setting.^{85,86,89} As shown in **Table 7**, in all three studies a technologist put the probes on the patients before the sleep recording. None of the publications commented on patient education, or the effects of it. Moreover, we note again that two of the three studies essentially compared the use of a type II monitor in the home versus the laboratory setting, and were only included as part of a best evidence approach.^{85,86} Overall, no conclusions can be drawn on this question for type II monitors.

Type III monitors

Table 9 summarizes the eight type III monitor studies in the home setting.^{31,32,54,90,96,97,100,101} As already noted in all papers studies were essentially unattended. A technologist attached the probes to the participants before they went home for the sleep study in three studies.^{31,90,97} The participants adjusted the electrodes themselves after a brief explanation/training by a technologist in two other studies.^{54,101} In Parra 1997,⁹⁶ probes were attached by a technologist in 50 people, and the remaining 39 participants hooked-up on their own at home (after receiving instructions). In Reichert 2003³² the participants attached the probes themselves without prior training; 24-hour telephone help line was available from the company, as well as video and detailed written instructions. There was no clear indication that concordance between the two measurements was better when technologist support was available. No mention on any effects of participant education on measurement concordance was made. However, we should note that sleep recordings with no interpretable data were excluded from the analyses in all papers. Overall, no robust conclusions can be drawn on this question for type III monitors.

Type IV monitors

Table 11 and **Table 12** summarize all type IV monitor studies conducted in settings other than a specialized sleep unit. As already noted in all papers studies were essentially unattended. In Schafer 1999,¹²³ a technologist attached the probes to the participants before they went home for the sleep study. In another study probe hook-up was randomly allocated to a technologist or the patient, but no differences were reported between these groups in terms of measurement concordance.¹¹¹ In five studies, the participants themselves performed the hook-up of the probes. No mention on any effects of participant education on measurement concordance was made.

Overall, there was no clear indication that concordance between the two measurements was better when technologist support was available. However, this may be attributed to the fact that sleep recordings with no interpretable data were excluded from the analyses in all papers. For example, Golpe 2002 had 7% data loss for technologist

hookup and 33% for patient hookup, and Bar 2003 had 3% “rejected studies” when the technologist did the setup in the sleep laboratory vs. 11% for patient setup. No robust conclusions can be drawn on this question for type IV monitors.

Synopsis for section B2a

For studies in the home setting, there is no direct data on whether and to what extent technologist support and patient education affect the comparison of portable monitors with facility-based polysomnography.

B3a. How do automated and manual scoring compare in the diagnosis of obstructive sleep apnea?

This question is addressed separately per type of monitor.

Facility-based polysomnography (Type I monitors)

Only two studies were identified that compared manual and automated scoring for facility-based PSG.^{46,51} Only facility-based PSG was used to assess AHI in 27 participants (Andreas 1993⁵¹) and 31 participants (Pittman 2004⁴⁶).

Andreas 1993⁵¹ compared manual scoring with “semi-automated” scoring. The software or its version is not mentioned. The software scored sleep stages using neurophysiological signals (i.e., the average frequency in the EEG, the presence of sleep spindles, K-complexes, EMG amplitude and other characteristics). It was “semi-automated” because a researcher had to provide input the (the minimum and maximum of EEG frequencies and EMG amplitudes) by consulting the polygraph prints. Overall, the mean difference in the number of respiratory events per hour was 8 more with the “semi-automated” scoring. The sensitivity and specificity of the “semi-automated” scoring to detect AHI>10 obstructive or mixed respiratory events/hour of sleep was 85% and 93% respectively.

Pittman 2004⁴⁶ compared automated scoring with the Morpheus I sleep scoring software with manual scoring by two experienced scorers. The automated scoring overestimated the manually derived AHI by 3 events/hour (95% limits of agreement: 17, -14) and 1 events/hour (95% limits of agreement: 17, -16), for the two scorers respectively. The area under the curve for the ability of automated scoring to predict more than 15 events/hour with manual scoring was very high (e.g., AUC=0.98 against the first scorer).

Type II monitors

None of the five type II monitor studies provided a direct comparison of manual and automated scoring methods.⁸⁵⁻⁸⁹ Two studies^{85,88} performed only automated scoring for the portable monitor, one used manual correction of automated scoring,⁸⁹ and two were manually scored.^{86,87} Thus, no conclusions can be drawn on this question for type II monitors.

Type III monitors

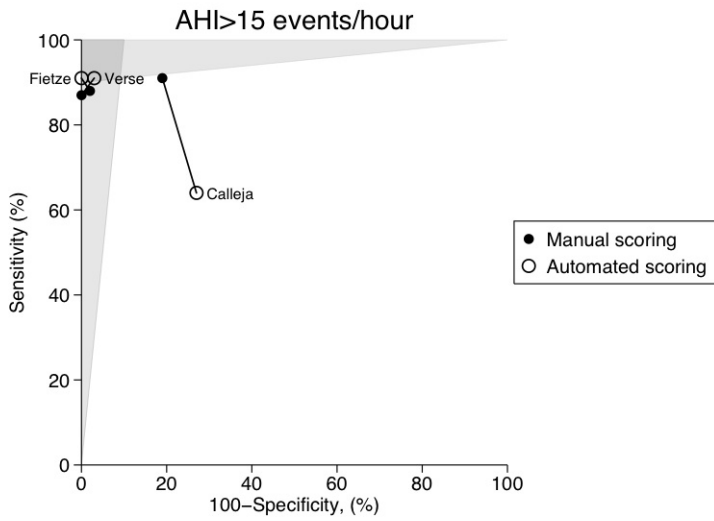
Dingli 2003⁵⁴ assessed the agreement of the Embletta monitor (unattended home setting) with facility-based PSG in the classification of subjects above or below an AHI of 15 events/hour. Agreement was good ($\kappa=0.62$) when the Embletta monitor recordings were scored manually, but was poor ($\kappa=0.10$) when automated scoring was used.

Carasco 1996 (type III monitor, unattended sleep study in a respiratory ward) described good concordance between the two monitors in classifying people above or below an AHI (or RDI for the portable monitor) of 20 events/hour. Using manual scoring, κ was 0.77; and using the best-performing algorithm for automated scoring, it was 0.60. The difference versus average analyses between the type III monitor and facility-based PSG yielded a systematic difference of -5.2 events/hour (95% limits of agreement: -31.4, 21.0) versus -2.9 events/hour (95% limits of agreement: -32.1, 26.3), for manual scoring and best-performing automated scoring algorithm (**Figure 9**). It should be noted that multiple other algorithms for automated scoring were assessed in that study, and we present results from the algorithm that had the best performance in the given set of recordings. Unless validated in an independent patient sample, the observed agreement with the best-performing algorithm cannot be taken at face value.

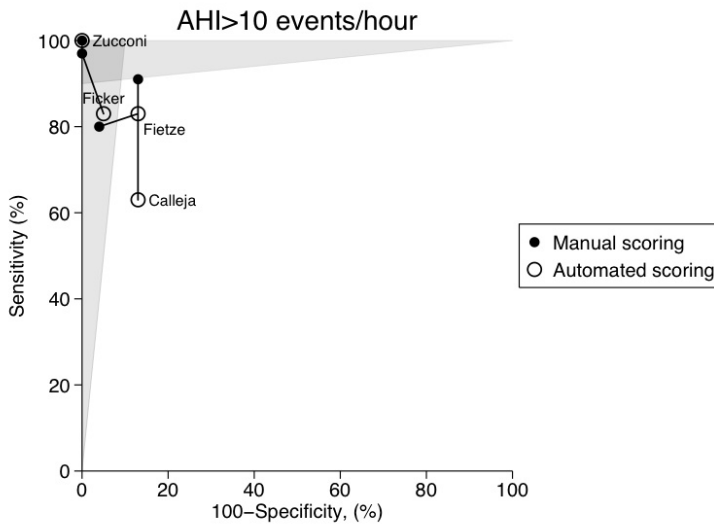
Finally, **Figure 28** shows how the sensitivity and specificity of the type III monitors depends on their scoring method. Studies that provided information with both scoring modalities using a cutoff of 15, 10 and 20 events/hour in facility-based PSG are depicted (all three studies were performed in the lab setting). As shown, differences between the automated and manual scorings may be large. Across the three panels of **Figure 28** manual scoring seems to perform better than automated scoring.

Figure 28. Diagnostic ability of type III monitors in different settings to identify AHI>15, >10 and >20 events/hour in laboratory-based polysomnography. Comparison of scoring methods.

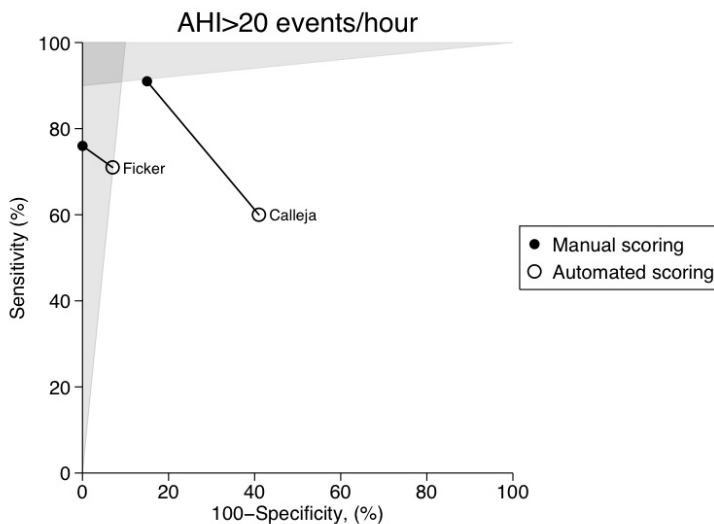
A



B



C



Lines connect estimates from the same study. Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded area have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both $LR+ > 10$ and $LR- < 0.1$. The figure depicts studies that used both automated and manual scoring for the type III monitor, and a cutoff of 15 events/h as suggestive of OSAHS in facility-based PSG.

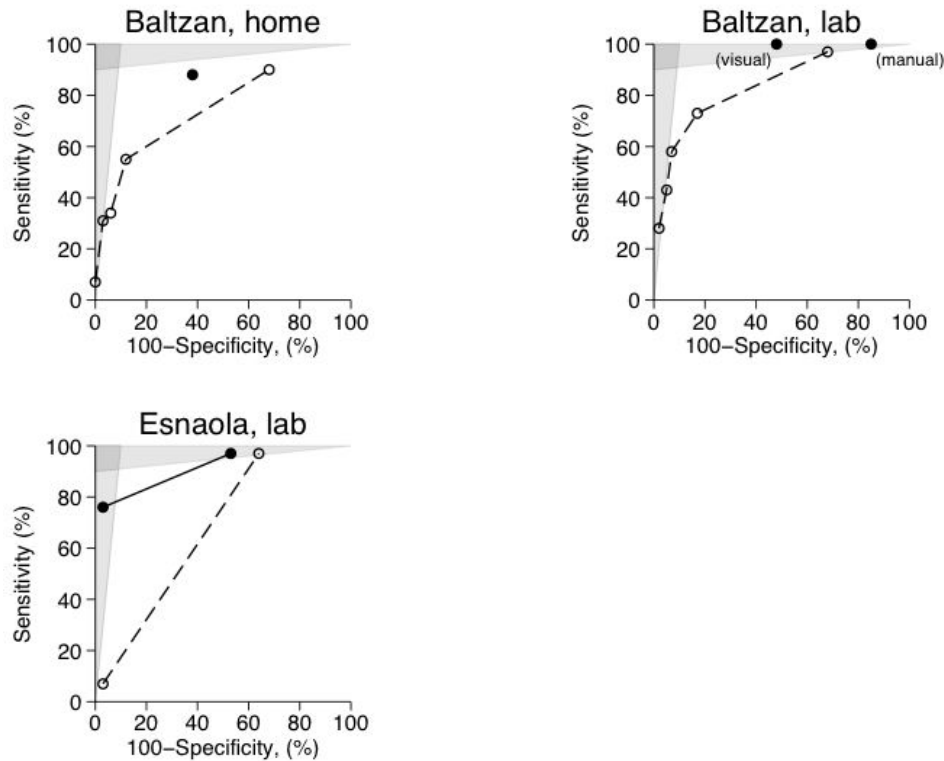
Type IV monitors

Four studies were eligible for this section. Three pertained to type IV monitors that recorded at least three bioparameters^{28,29,111} and one to type IV monitors recording two bioparameters (Baltzan 2000⁸¹)

Esnaola 1996²⁸ found an average difference of 2 events/hour (95% limits of agreement: -35, 38) between facility-based PSG and manual scoring of type IV monitor. The corresponding average bias in the measurements was 9 events/hour (95% limits of agreement: -42, 61) with automated scoring (**Figure 21**). Similarly, the intraclass correlation coefficients with the facility-based PSG measurements were higher with manual scoring (0.72 versus 0.47, respectively).²⁸

Figure 29 juxtaposes manual and automated scoring in the square plot for two studies^{28,81} that used both scoring methods to predict AHI >15 events/hour of sleep in facility-based PSG. Manual scoring generally had better diagnostic ability. The same was true in Koziej 1994²⁹ and Esnaola 1996²⁸ for a cutoff of 10 events/hour of sleep in facility-based PSG. Finally, Golpe 2002¹¹¹ found similarly high areas under the curve for manual and automated scoring (0.89 and 0.86, respectively) using a cutoff of 10 events/hour of sleep in facility-based PSG.

Figure 29. Diagnostic ability of type IV monitors in different settings to identify AHI>15 events/hour in laboratory-based polysomnography. Comparison of scoring methods.



Dashed or solid lines connect sensitivity/specificity pairs that have been estimates with different thresholds for the type IV monitor. These lines do not represent the ROC curves from the pertinent studies. Studies lying on the left shaded area have a positive likelihood ratio of 10 or more. Studies lying on the top shaded area have a negative likelihood ratio of 0.1 or less. Studies lying on the intersection of the grey areas (darker grey polygon) have both LR+>10 and LR-<0.1.

Filled dots represent manual scoring, and empty circles automated scoring. Baltzan in the lab setting used a visual categorization in suggestive of OSAHS or not (visual) and manual counting of the respiratory events (manual).

The figure depicts studies that used both automated and manual scoring for the type IV monitor, and a cutoff of 15 events/h as suggestive of OSAHS in facility-based PSG.

Synopsis for section B3a

Overall, manual scoring or manual editing of automated scoring seems to have better agreement with facility-based PSG compared to automated scoring in the studies that assessed both scoring methods. We note that the automated scoring algorithms from different monitors are different, and their ability to recognize respiratory events may vary with the specific software version or settings. Therefore, this observation may not be generalizable to all monitors.

B3b. What errors related to automated and manual scoring are reported?

Overall there were no specific errors attributed to manual or automated scoring methods per se in the assessed studies. Carasco 1996⁵³ noted that the automated scoring was discordant from the manual scoring because the identification of hypopneas was problematic; in contrast apneas were correctly recognized by the software.

Synopsis for section B3b

We did not identify detailed data on the specific types of errors that are related to automated or manual scoring rather than other parameters. No robust conclusion can be given for this question.

B4. For studies of portable monitoring in the home setting, what errors related to unattended use are reported?

This question is addressed separately per type of monitor.

Type II monitors

Three studies of type II monitors were performed in the home setting.^{85,86,89} As shown in **Table 7**, all three studies were unattended. None of the publications linked any errors directly to unattended use. However, two of the three studies essentially compared the use of a type II monitor in the home versus the hospital or sleep laboratory setting (they were included as part of a best evidence approach^{85,86}). It can be argued that an explanation of any differences in errors is at least related to unattended use at home.

In Gagnadoux 2002,⁸⁵ 11.1% of 111 attended hospital-based studies (95% confidence interval: 4.9, 17.4%) did not have interpretable signals for more than 180 minutes in a minimum set of signals (at least one EEG channel, one EOG, EMG, airflow, thoracoabdominal movements, and oximetry). The corresponding proportion was 23.4% (95% confidence interval: 19.2, 27.7) in the unattended home studies, higher beyond chance. In Iber 2004,⁸⁶ technologists made belt adjustments 60 times in 41 studies in laboratory-based PSG. However, no differences were observed in signal quality scores across the 33 out of 64 cases for which paired recordings in all channels were available for the attended and unattended sleep studies. Portier 2000⁸⁹ noted that higher discrepancies were associated with poor quality airflow recordings at home (unattended).

Finally, Kapur 2000⁴⁹ studied the rates of sensor loss in unattended home sleep monitoring with type II monitors (Compumedics PS-2) in the context of the Sleep Heart Health Study (SHHS). The participants were studied only with the type II monitor. We report this study here because it may be hypothesized that the errors or signal loss in this study are relevant to it being unattended at home. Approximately 91% of 6802 participants had a successful first study, and 4% needed more than one additional attempt before obtaining a successful recording. In the end 5% of people had unsuccessful studies. On average, the thoracoabdominal effort channels and the airflow channels yielded valid recordings 79% of the total recording time; chin EMG 81% of the time; EEG 89% of the time; EOG 93% of the time and oximetry 93% of the total recording time. Table 15 shows significant predictors of signal duration from this study, and how they affect the different signals.

Table 15. Significant predictors of valid signal duration in unattended home monitors (type II) in the Sleep Heart Health Study

Signal	Male gender	Increasing BMI	Increasing age	Increasing RDI
Overall		↓		
Respiratory		↓		↑
Staging	↓			
Effort-Abdominal		↓	↓	
Effort-Thoracic		↓		
Chin EMG	↓			

BMI: Body mass index; EMG: Electromyogram; RDI: respiratory distress index

Type III monitors

None of the eight studies in the home setting^{31,32,54,90,96,97,100,101} was attended. However, Reichert 2003³² used a portable monitor that has the ability to alert the evaluated person if a probe is not correctly connected or does not give a valid signal (NovaSom QSG). The monitor was used for three nights in the home setting. Three out of 51 recordings were lost because of memory chip malfunction (another three subjects did not use the portable machine at all). None of the 45 people who successfully used the monitor reported having difficulties in using the machine. There were no invalid home studies due to misapplied probes and no calls made to the 24-hour help line that was available. However, there are no data on how many times the portable monitor alerted the patients to correct dislodged probes during the home study. Moreover, there were no comparisons with activated versus non-activated alert function of the NovaSom QSG monitor. Thus, it cannot be assessed whether the built-in alerting system prevented data losses or data corruption.

Finally, in the White 1995 study³¹ recorded data were sent to a technologist via modem at least every 30 minutes. After inspecting the downloaded data, the technologist would call the participants if corrections were necessary. Half of the participants (53%) were called during the night to ensure good recordings. On the average 0.8 calls per participant were needed.

Type IV monitors

All type IV monitor studies in the home setting were unattended. Thus, no conclusions can be drawn on this question.

Synopsis for section B4

No studies of type II monitors associated directly any specific errors with unattended use. However, signal loss was more often observed for home studies, and one study associated discrepancies in the AHI measurement with poor quality airflow signals in the unattended home-based recordings.

Two studies of type III monitors in the home setting were unattended but used a built in alert to notify the user for errors, or a modem connection to send recorded data to a technologist. Indirect evidence suggests that errors were possibly prevented using the aforementioned mechanisms (instead of physical presence).

B5. Do the reported complications, harms, and adverse events differ for portable multichannel sleep testing and facility-based polysomnography?

FDA data on adverse events of medical devices used in sleep studies

Several adverse events and complications were reported to the FDA. It is not possible to estimate the prevalence of the adverse events using this source, because the total number of sleep recordings (exposure, denominator) is unknown. Overall, the reported adverse events pertained to burns, possible allergic reactions, and eye irritations. Most of them were reported to be mild or not extensive. Complications secondary to device malfunction and wrong or undocumented device use were reported.

Specifically, burning, tingling, or shocking sensations from probes in various sites were reported (neck, nose and lips, forehead, calves). Responsible electrodes were grounding electrodes, electrodes for EEG/EMG, or a thermistor probe whose covering/insulation was damaged and the soldering joint was exposed. Electrical burns of various degrees (small first, second, and third degree burns) were reported in the cheeks and on the nose and the lips secondary to thermistor malfunction. Thermal burns were reported on the abdomen after a technologist erroneously plugged the strain gauges to the battery outlet of the direct current converter instead of the input jack. A patient suffered a thermal burn in the finger after touching an overheated battery. Chemical burns secondary to the use of disinfectants on masks (e.g., use of “cidex opa” resulted in irritations on the face) or conductives for ECG or EMG electrodes were also reported. Some burns resulted in loss of hair at the site of the electrode or skin discolorations in dark skinned patients. Chemical irritation or allergic reaction to the adhesive tape that was used to hold a digit oximeter was also reported. Finally, it was reported that patients had eye irritations after a sleep study, presumably because conductive material got in their eyes during showering (long term damage was claimed in at least one case).

Laboratory-based polysomnography (Type I monitors)

A very large prospective study¹³⁸ on the safety of facility-based PSG used data on 16,084 sleep recordings conducted in 17 USA centers between January 1, 2002 and June 30, 2003. One death was reported two weeks after the facility-based PSG study in a 60 year old patient with known coronary artery disease and dilated cardiomyopathy, who had an ejection fraction <20%. Hence, the prevalence of death reported by the study was 0.0062% (95% confidence interval: 0, 0.019%). There were 28 events prompting immediate attention (0.17% [95% confidence interval: 0, 0.24%]). These pertained

mostly to arrhythmias associated with sleep disordered breathing (shortness of breath and chest pain occurred in one patient). Finally, there were 28 potentially alarming additional events that were noted by members of the scoring team who were reviewing the patient charts. These were mainly complex ventricular arrhythmias. Overall, the study reported that any complication occurred in 0.35% of sleep studies (95% confidence interval: 0.26, 0.45%).

Type II monitors

No mention on potential complications, harms or adverse events was made in the five studies of type II monitors.⁸⁵⁻⁸⁹

Type III monitors

No mention on potential complications, harms or adverse events was made in the 22 studies that assessed type III monitors.^{31,32,50,52-57,90-102}

Type IV monitors

No mention on potential complications, harms or adverse events was made in the 46 studies that assessed type IV monitors.^{26-30,42,58,59,76,80-84,103-134}

Synopsis for section B5

The rate and severity of adverse events in sleep studies are low. In a large study, complications were identified in less than 0.5% of the recordings during facility-based PSG. Complications did not arise in the remaining studies, and mostly minor harms were reported to the FDA adverse events database. This conclusion applies to both facility-based PSG and to portable monitors. An advantage of facility-based PSG was the ability of specialized personnel to intervene in case of events necessitating immediate attention (i.e., mostly arrhythmias associated with sleep disordered breathing).

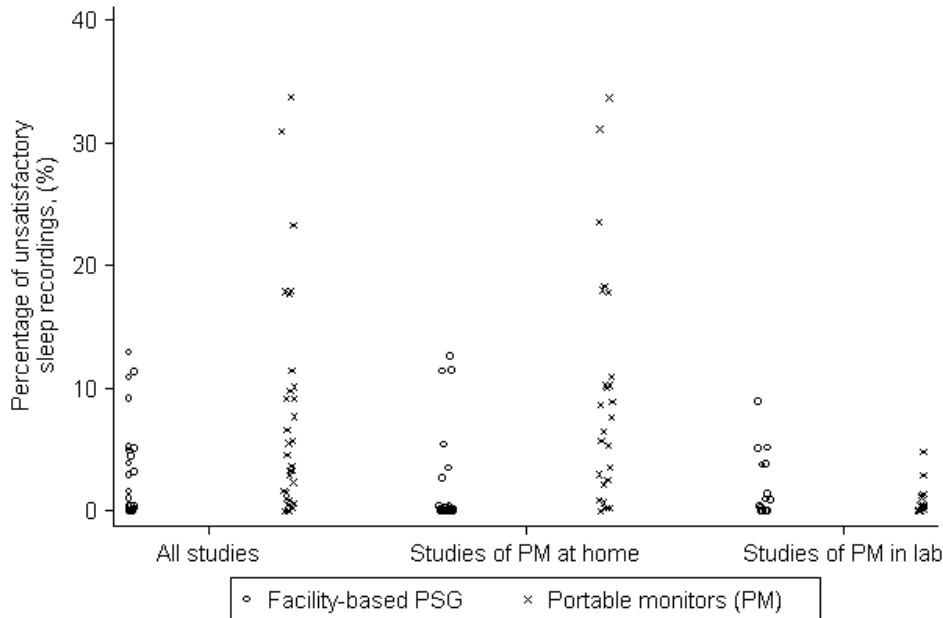
B6. Do rates or types of data loss and data corruption differ for portable multichannel sleep testing and facility-based polysomnography?

In addition to studies that were eligible for other key questions, we identified 12 prospective and non-overlapping studies (each including more than 100 people) that described rates of unsatisfactory sleep recordings during facility-based PSG or during sleep studies with portable monitors.^{46,49,139-148} All types of data loss and unsatisfactory and non-analyzable sleep recordings were according to each study's definitions: data loss secondary to user errors, bad signal quality, malfunctions or incorrect probe hook-up either in facility-based PSG or in studies using portable monitors.

The unit of analysis is the number of sleep recordings that were performed rather than the number of participants. Reporting was very inconsistent. Most papers either did not report rates of data loss at all, or indicated that all included sleep recordings were "satisfactory", without mentioning whether any unsatisfactory studies were repeated or excluded upfront. Overall, 36 papers clearly reported non-zero proportion of unsatisfactory recordings either for facility-based PSG or for portable monitors. **Figure 30** plots the percentages of lost data overall and per setting of the portable monitors for these studies. As evident from the graph, portable monitors have higher rates and variability of data loss compared to facility-based PSG. This is mainly driven by the

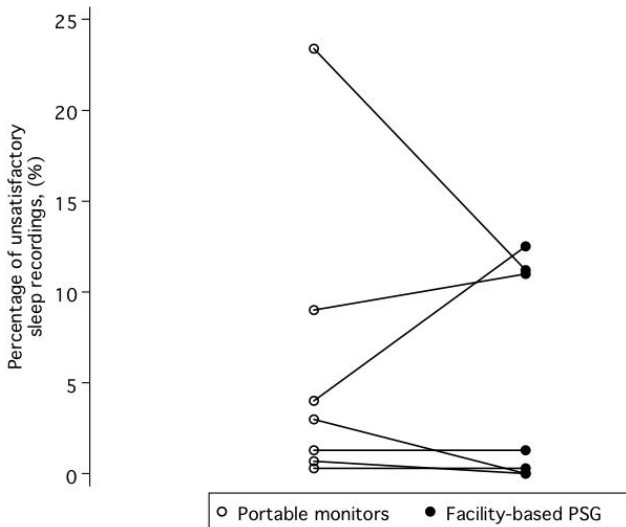
higher percentage of unsatisfactory recordings with portable monitors in the home setting (**Figure 30**). **Figure 31** depicts studies that clearly reported differential rates of unsatisfactory sleep recordings with different monitors.

Figure 30. Rates of unsatisfactory sleep recordings for facility-based PSG and portable monitors in studies that clearly reported the pertinent information.



Shown are rates of unsatisfactory studies secondary to any reason for facility-based PSG and portable monitors, in studies that clearly reported the pertinent information. If the number of unsatisfactory recordings was reported only for one type of monitor (e.g., facility-based PSG) and was unknown for the other monitor (e.g., portable monitor; when applicable), 0% loss was assumed for the unknown rate of unsatisfactory recordings. A small amount of random noise has been added to the plotted points to facilitate readability.

Figure 31. Comparative rates of unsatisfactory sleep recordings for facility-based polysomnography and portable monitors.



Shown are rates of unsatisfactory studies secondary to any reason for facility-based PSG and portable monitors, in studies that clearly reported the pertinent information for both categories of monitors. In all except for one case the portable monitor was used at home; the exception is a study of portable monitor in the laboratory setting where the rate of unsatisfactory recordings was 1.3% with both monitors.

The most commonly reported reasons for unsatisfactory recordings in facility-based PSG were technical reasons (without details); loss or poor quality of at least one signal (oximetry was often mentioned, along with airflow, EEG and EMG signals); and very short duration of sleep (usually less than 180 minutes). For portable monitors, the most common reasons were related to the failure of the user to operate the device correctly (probes plugged in the wrong inputs, failure to switch the machine on, failure to connect probe to the machine); poor signal quality (often secondary to probe detachment; oximetry was most often mentioned, as it is almost always recorded, followed by airflow and effort signals); recording loss because of errors during data downloading or faulty memory chip; short sleep duration (typically less than 180 minutes); and battery failure.

Synopsis for section B6

Rates of unsatisfactory studies and data corruption are higher for portable monitors in the home setting, compared to facility-based PSG, or portable monitors in the sleep laboratory setting. This is frequently attributed to user errors during device operation or probe hook-up for home studies, or also probably secondary to the absence of an attending technologist.

Overview

The following overview summarizes our findings to the key questions.

Note that generally, the included studies evaluated populations with high likelihood for OSAHS, without co-morbidities that may affect sleep parameters (e.g. cardiac insufficiency), who were young (approximately 50 years old), predominantly male, and with high BMI. The technology assessment excluded people who had been operated for OSAHS. The aforementioned caveats should be taken in mind when assessing the applicability of the findings to the Medicare population (see below for a relevant discussion).

Key Question A1.

Is laboratory-based polysomnography considered a “gold standard” for the diagnosis of sleep apnea?

Facility-based PSG is the reference method to identify people with *AHI suggestive of OSAHS*.

This does not mean that facility-based PSG is an error-free “gold standard” for the *diagnosis of OSAHS*. The diagnosis of OSAHS typically requires additional information (symptoms and signs, differentiation of other conditions that affect sleep).

Furthermore, there is evidence that for people with a high probability for OSAHS, use of facility-based PSG does not result in better outcomes over an ambulatory approach in terms of diagnosis and CPAP titration.

Key Question A2.

Which are the appropriate methods for the comparison of diagnostic test performance for obstructive sleep apnea?

Ideally, one would like to evaluate hard outcomes in subjects who were managed based on information from different sleep monitors. Typically, hard outcomes include mortality, cardiovascular morbidity, incidence of accidents, and similar endpoints. A randomized controlled trial (RCT) would be the ideal study design.

The vast majority of the existing studies directly compares the measurements of AHI (or RDI, or related indices) obtained with the portable monitors or facility-based PSG. When two measurements are compared, one may:

1. Assess the agreement of the individual measurements directly. This answers the question of whether the two measurements (e.g., AHI from facility-based PSG and AHI or RDI from portable monitors) are *potentially interchangeable*.
 - a. Difference versus average analyses (Bland-Altman plots) are a suitable method to address this question
2. Assess the ability of portable monitors to predict AHI measurements in facility-based PSG that are *suggestive of OSAHS* or *not suggestive of OSAHS*.
 - a. Assuming that the reference standard of facility-based PSG has negligible error in the measurement of AHI, sensitivity, specificity and likelihood-ratio-based calculations are suitable.

The philosophy behind these questions is very different. This is also true for the assumptions underlying the statistical methods that are used in their analyses. It is very common for researchers to use wrong statistical methods (Pearson's correlation coefficient and ordinary least squares regressions) in measurement comparison studies.

Key Question A3.

Classification of portable sleeping monitoring devices (Types I, II, III and IV).

The American Sleep Disorders Association (ASDA) classified the devices that have been used in sleep studies into four categories, based on the signals they record. For example, facility-based PSG is considered a type I device, and is the most comprehensive. There is a progressive loss of sleep parameter information from type I and II monitors to type IV monitors. Information on sleep staging is lost in Type III monitors. Some or all information on airflow /effort is lost in Type IV monitors (Type III devices have to have at two airflow channels one of which can be an effort channel).

The ASDA classification serves to characterize the sleep monitors for operational purposes. It does not imply that e.g., all Type IV monitors are equivalent.

There are some difficulties with the aforementioned classification scheme, which cannot explicitly classify all existing portable monitors. This is especially true for newer portable monitors that measure bioparameters proposed in later years (and are not listed in the ASDA classification scheme). For example, according to the ASDA scheme, Type IV monitors measure only one or two bioparameters. However, there are monitors that measure more than two bioparameters but they do not meet the criteria for Type III devices. These would remain "unclassified". In this technology assessment, we broadened the definition of Type IV to include all monitors that *do not meet the criteria for Type III devices* (despite the fact that they record more than two bioparameters). However we report separately Type IV monitors that record only one or two bioparameters from those that record more than two.

Key Question B1.

Ability of facility-based polysomnography versus portable monitors to predict response to CPAP and changes in clinical outcomes after CPAP treatment.

Baseline AHI from facility-based PSG is only modestly associated with response to CPAP (as conveyed by the assessed outcomes) among people with high probability for severe OSAHS. Thus, differences in baseline AHI cannot be used to accurately predict CPAP use or response to CPAP in this population.

In addition, a randomized study suggested that the increased accuracy in AHI estimation with facility-based PSG versus a type IV monitor does not translate to more accurate prediction of response to CPAP with respect to a quality of life outcome. This finding is in line with the aforementioned caveats and pertains to patients who are at the severe end of the AHI spectrum

Key Question B2.

How does the performance of portable monitors compare with facility-based polysomnography for the diagnosis of OSAHS?

The AHI (or RDI) measurements from portable monitors and facility-based PSG are not interchangeable (especially in the higher end of the AHI spectrum):

Difference versus average analyses suggest that substantial differences in the AHI may be encountered between type II monitors and facility-based PSG, especially when the studies are not performed simultaneously, but at different nights. Even larger differences compared with facility-based PSG cannot be excluded for type III monitors; and more so for type IV monitors. This was true both for manual and automated scoring.

Nevertheless, portable monitors may be able to predict AHI suggestive of OSAHS in facility-based PSG. This is compatible with the above analyses, when the monitors disagree in how large a large AHI value is, but agree that it is high enough to be “suggestive of OSAHS”:

Based on limited data, type II monitors may identify people with AHI suggestive of OSAHS with high positive likelihood ratios and low negative likelihood ratios. Type III monitors may have the ability to predict AHI suggestive of OSAHS with high positive likelihood ratios and low negative likelihood ratios for various AHI cutoffs in laboratory-based PSG, especially when manual scoring is employed. Studies of type IV monitors that record at least three bioparameters showed high positive likelihood ratios and low negative likelihood ratios. Studies of type IV monitors that record one or two bioparameters also had high positive likelihood ratios and low negative likelihood ratios, at least for selected sensitivity and specificity pairs from ROC curve analyses.

Overall, the ability of portable monitors to predict AHI with facility-based PSG appears to be worse in studies conducted in the home setting compared to studies in the specialized sleep laboratory. Between-night variability is a plausible explanation: in the sleep-lab setting measurements are simultaneous, whereas in other settings (home, hospital ward, outside the sleep lab) measurements are performed in different nights.

What does AHI suggestive of OSAHS mean?

The actual diagnosis of OSAHS is not set solely on the basis of high AHI values. Whether an AHI value “suggestive of OSAHS” reflects the true diagnosis of OSAHS or not is dependent on several factors:

1. The actual AHI cutoff that was used.
2. The presence of conditions that affect sleep quality and may not equally identified by facility-based PSG and some portable monitors:
 - a. People with cardiac insufficiency or atrial flutter who have Cheyne-Stokes respiration, and people with respiratory disturbances such as chronic obstructive pulmonary disease, obesity hypoventilation syndrome, or people with other sleep disorders such as narcolepsy.
 - b. Periodic Limb Movements in Sleep (PLMS), which are observed in the majority of patients with the Restless Leg Syndrome (RLS).
3. The prior probability of OSAHS. People with high prior probability of OSAHS would be:

- a. Referrals from physicians familiar with OSAHS
- b. People with symptoms or suggestive clinical profile

Applicability to the Medicare population:

Three caveats may be made:

First, the prior probability of OSAHS among Medicare beneficiaries may be lower than the prior probability of OSAHS among subjects analyzed in the included studies:

1. The average age of the participants in the analyzed studies was recorded. The median value was 50 to 52 years. Moreover, the majority of the subjects were males and they had BMI above 25 kg/m². Finally, in most studies subjects were referred for PSG by sleep or respiratory physicians. On the other hand, Medicare beneficiaries are older (≥ 65 years), are not predominantly male, and may often have comorbidities that affect sleep quality.
2. There is evidence that, in the elderly, obesity and daytime somnolence may not be as strongly associated with OSAHS as in middle aged people.¹⁵ In other words, a high AHI or RDI value would not be as strongly associated with an *actual diagnosis of OSAHS*.

Second, facility-based PSG can differentiate OSAHS from other conditions that cause sleep disturbances (such as PLMS/RLS, or conditions associated with mixed or central apneas or hypopneas). The following caveats may be made:

1. Conditions that are associated with sleep disturbances with a central component (such as e.g., cardiac insufficiency) are prevalent among Medicare beneficiaries. This may be true for respiratory conditions (chronic obstructive pulmonary disease, obesity hypoventilation syndrome and other hypoventilatory syndromes) as well as other sleep disorders like narcolepsy. This caveat probably pertains to type III and type IV monitors.
2. Oximeters (and some monitors that do not assess airflow or effort) may not be able to differentiate PLMS/RLS from OSAHS. The prevalence of RLS is increased among older adults (up to 8% in people older than 60 years, compared to approximately 5% in people who are in their forties).¹³⁷ Moreover, PLMS/RLS is more common in women,¹³⁷ who were the minority in the studied populations. This caveat pertains to some type IV monitors that assess one or two bioparameters only.

Finally, in almost all eligible studies the investigators were associated with sleep clinics and/or sleep laboratories. Study investigators are probably very familiar with the diagnosis and treatment of sleep disorders (including ones other than OSAHS) even when this is not clearly stated in the primary papers). One cannot necessarily extrapolate the findings in these studies to circumstances where health care providers with less training and experience might use the devices.

For studies in the home setting, do technologist support and patient education affect the comparison of portable and with facility-based polysomnography for the diagnosis of OSAHS?

For studies in the home setting, there is no direct data on whether and to what extent technologist support and patient education affect the comparison of portable monitors with facility-based polysomnography.

Key Question B3.

How do automated and manual scoring compare in the diagnosis of obstructive sleep apnea?

Overall, manual scoring or manual editing of automated scoring seems to have better agreement with facility-based PSG compared to manual scoring in the studies that assessed both scoring methods. We note that the automated scoring algorithms from different monitors are different, and their ability to recognize respiratory events may vary with the specific software version or settings.

What errors related to automated and manual scoring are reported?

We did not identify detailed data on the specific types of errors that are related to automated or manual scoring rather than other parameters. No robust conclusion can be given for this question.

Key Question B4.

For studies of portable monitoring in the home setting, what errors related to unattended use are reported?

No studies of type II monitors associated directly any specific errors with unattended use. However, signal loss was more often observed for home studies. Two studies of type III monitors in the home setting were unattended but used a built in alert to notify the user for errors, or a modem connection to send recorded data to a technologist. Indirect evidence suggests that errors were possibly prevented using the aforementioned mechanisms (instead of physical presence).

Key Question B5.

Do the reported complications, harms, and adverse events differ for portable multichannel sleep testing and facility-based polysomnography?

The rate and severity of adverse events in sleep studies are low. In a large study, complications were identified in less than 0.5% of the recordings during facility-based PSG. Complications were not reported in the remaining studies, and mostly minor harms were reported to the FDA adverse events database. This conclusion applies to both facility-based PSG and to portable monitors. An advantage of facility-based PSG was the ability of specialized personnel to intervene in case of events necessitating immediate attention (i.e., mostly arrhythmias associated with sleep disordered breathing).

Key Question B6.

Do rates or types of data loss and data corruption differ for portable multichannel sleep testing and facility-based polysomnography?

Rates of unsatisfactory studies and data corruption are higher for portable monitors in the home setting, compared to facility-based PSG, or portable monitors in the sleep laboratory setting. This is probably attributed to user errors during device operation or probe hook-up for home studies, and probably also to the absence of an attending technologist.

References and included studies

1. Practice parameters for the indications for polysomnography and related procedures. Polysomnography Task Force, American Sleep Disorders Association Standards of Practice Committee. *Sleep* 1997; 20(6):406-422.
2. Marin JM, Carrizo SJ, Vicente E, Agusti AG. Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study. *Lancet* 2005; 365(9464):1046-1053.
3. Yaggi HK, Concato J, Kernan WN, et al. Obstructive sleep apnea as a risk factor for stroke and death. *N Engl J Med* 2005; 353(19):2034-2041.
4. Young T, Palta M, Dempsey J, et al. The occurrence of sleep-disordered breathing among middle-aged adults. *N Engl J Med* 1993; 328(17):1230-1235.
5. Shahar E, Whitney CW, Redline S, et al. Sleep-disordered breathing and cardiovascular disease: cross-sectional results of the Sleep Heart Health Study. *Am J Respir Crit Care Med* 2001; 163(1):19-25.
6. Nieto FJ, Young TB, Lind BK, et al. Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study. Sleep Heart Health Study. *JAMA* 2000; 283(14):1829-1836.
7. Peppard PE, Young T, Palta M, Skatrud J. Prospective study of the association between sleep-disordered breathing and hypertension. *N Engl J Med* 2000; 342(19):1378-1384.
8. Young T, Peppard P, Palta M, et al. Population-based study of sleep-disordered breathing as a risk factor for hypertension. *Arch Intern Med* 1997; 157(15):1746-1752.
9. Babu AR, Herdegen J, Fogelfeld L, Shott S, Mazzone T. Type 2 diabetes, glycemic control, and continuous positive airway pressure in obstructive sleep apnea. *Arch Intern Med* 2005; 165(4):447-452.
10. Gruber A, Horwood F, Sithole J, Ali NJ, Idris I. Obstructive sleep apnoea is independently associated with the metabolic syndrome but not insulin resistance state. *Cardiovasc Diabetol* 2006; 5:22.
11. Howard ME, Desai AV, Grunstein RR, et al. Sleepiness, sleep-disordered breathing, and accident risk factors in commercial vehicle drivers. *Am J Respir Crit Care Med* 2004; 170(9):1014-1021.
12. Teran-Santos J, Jimenez-Gomez A, Cordero-Guevara J. The association between sleep apnea and the risk of traffic accidents. Cooperative Group Burgos-Santander. *N Engl J Med* 1999; 340(11):847-851.
13. 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(4):1543-1579.
14. Young T, Peppard PE, Taheri S. Excess weight and sleep-disordered breathing. *J Appl Physiol* 2005; 99(4):1592-1599.
15. 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(8):893-900.
16. Ancoli-Israel S, Kripke DF, Klauber MR, et al. Sleep-disordered breathing in community-dwelling elderly. *Sleep* 1991; 14(6):486-495.
17. Bixler EO, Vgontzas AN, Ten HT, Tyson K, Kales A. Effects of age on sleep apnea in men: I. Prevalence and

- severity. *Am J Respir Crit Care Med* 1998; 157(1):144-148.
18. Bixler EO, Vgontzas AN, Lin HM, et al. Prevalence of sleep-disordered breathing in women: effects of gender. *Am J Respir Crit Care Med* 2001; 163(3 Pt 1):608-613.
 19. Duran J, Esnaola S, Rubio R, Iztueta A. Obstructive sleep apnea-hypopnea and related clinical features in a population-based sample of subjects aged 30 to 70 yr. *Am J Respir Crit Care Med* 2001; 163(3 Pt 1):685-689.
 20. Young T, Peppard PE, Gottlieb DJ. Epidemiology of obstructive sleep apnea: a population health perspective. *Am J Respir Crit Care Med* 2002; 165(9):1217-1239.
 21. Ferber R, Millman R, Coppola M, et al. Portable recording in the assessment of obstructive sleep apnea. *ASDA standards of practice. Sleep* 1994; 17(4):378-392.
 22. Lux L, Boehlecke B, Lohr KN. Effects of portable monitoring devices for diagnosing obstructive sleep apnea: Update of a systematic review. 290-20-0016. 2006. Agency for Healthcare Research and Quality.
 23. Ayappa I, Norman RG, Krieger AC, et al. Non-Invasive detection of respiratory effort-related arousals (REras) by a nasal cannula/pressure transducer system. *Sleep* 2000; 23(6):763-771.
 24. Masa JF, Corral J, Martin MJ, et al. Assessment of thoracoabdominal bands to detect respiratory effort-related arousal. *European Respiratory Journal* 2003; 22(4):661-667.
 25. Rechtschaffen A, Kales A. A manual of standardized techniques and scoring system for sleep staging of human subjects. Los Angeles: Brain Information Service and Brain Research Institute, 1968.
 26. Fleury B, Rakotonanahary D, Hausser-Hauw C, Lebeau B, Guilleminault C. A laboratory validation study of the diagnostic mode of the Autoset system for sleep-related respiratory disorders.[erratum appears in *Sleep* 1996 Sep;19(7):601]. *Sleep* 1996; 19(6):502-505.
 27. Pepin JL, Levy P, Lepaulle B, Brambilla C, Guilleminault C. Does oximetry contribute to the detection of apneic events? Mathematical processing of the SaO₂ signal. *Chest* 1991; 99(5):1151-1157.
 28. Esnaola S, Duran J, Infante-Rivard C, Rubio R, Fernandez A. Diagnostic accuracy of a portable recording device (MESAM IV) in suspected obstructive sleep apnoea. *European Respiratory Journal* 1996; 9(12):2597-2605.
 29. Koziej M, Cieslicki JK, Gorzelak K, Sliwinski P, Zielinski J. Hand-scoring of MESAM 4 recordings is more accurate than automatic analysis in screening for obstructive sleep apnoea. *European Respiratory Journal* 1994; 7(10):1771-1775.
 30. 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; 4(3):207-212.
 31. White DP, Gibb TJ, Wall JM, Westbrook PR. Assessment of accuracy and analysis time of a novel device to monitor sleep and breathing in the home.[see comment]. *Sleep* 1995; 18(2):115-126.
 32. Reichert JA, Bloch DA, Cundiff E, Votteri BA. Comparison of the NovaSom QSG, a new sleep apnea home-diagnostic system, and polysomnography. *Sleep Medicine* 2003; 4(3):213-218.
 33. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1(8476):307-310.

34. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet* 1995; 346(8982):1085-1087.
35. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989; 45(1):255-268.
36. Lin LI. A note on the concordance correlation coefficient. *Biometrics* 2000; 56:324-325.
37. Bradley EL, Blackwood LG. Comparing paired data: a simultaneous test for means and variances. *Am Statistician* 1989; 43:234-235.
38. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994; 271(9):703-707.
39. Flemons WW, Littner MR. Measuring agreement between diagnostic devices. *Chest* 2003; 124(4):1535-1542.
40. The AASM manual for the scoring of sleep and associated events. Rules, terminology, and technical specifications. Westchester, IL: American Academy of Sleep Medicine., 2007.
41. 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(5):667-689.
42. Wiltshire N, Kendrick AH, Catterall JR. Home oximetry studies for diagnosis of sleep apnea/hypopnea syndrome: limitation of memory storage capabilities. *Chest* 2001; 120(2):384-389.
43. Zafar S, Ayappa I, Norman RG, et al. Choice of oximeter affects apnea-hypopnea index. *Chest* 2005; 127(1):80-88.
44. Norman RG, Pal I, Stewart C, Walsleben JA, Rapoport DM. Interobserver agreement among sleep scorers from different centers in a large dataset. *Sleep* 2000; 23(7):901-908.
45. Collop NA. Scoring variability between polysomnography technologists in different sleep laboratories. *Sleep Med* 2002; 3(1):43-47.
46. Pittman SD, MacDonald MM, Fogel RB, et al. Assessment of automated scoring of polysomnographic recordings in a population with suspected sleep-disordered breathing. *Sleep* 2004; 27(7):1394-1403.
47. Schaltenbrand N, Lengelle R, Toussaint M, et al. Sleep stage scoring using the neural network model: comparison between visual and automatic analysis in normal subjects and patients. *Sleep* 1996; 19(1):26-35.
48. Whitney CW, Gottlieb DJ, Redline S, et al. Reliability of scoring respiratory disturbance indices and sleep staging. *Sleep* 1998; 21(7):749-757.
49. Kapur VK, Rapoport DM, Sanders MH, et al. Rates of sensor loss in unattended home polysomnography: the influence of age, gender, obesity, and sleep-disordered breathing. *Sleep* 2000; 23(5):682-688.
50. Redline S, Tosteson T, Boucher MA, Millman RP. Measurement of sleep-related breathing disturbances in epidemiologic studies. Assessment of the validity and reproducibility of a portable monitoring device. *Chest* 1991; 100(5):1281-1286.
51. Andreas S, von BB, Magnusson K, Kreuzer H. Validation of automated sleep stage and apnoea analysis in suspected obstructive sleep apnoea. *European Respiratory Journal* 1993; 6(1):48-52.

52. Calleja JM, Esnaola S, Rubio R, Duran J. Comparison of a cardiorespiratory device versus polysomnography for diagnosis of sleep apnoea. *European Respiratory Journal* 2002; 20(6):1505-1510.
53. Carrasco O, Montserrat JM, Lloberes P, et al. Visual and different automatic scoring profiles of respiratory variables in the diagnosis of sleep apnoea-hypopnoea syndrome. *European Respiratory Journal* 1996; 9(1):125-130.
54. Dingli K, Coleman EL, Vennelle M, et al. Evaluation of a portable device for diagnosing the sleep apnoea/hypopnoea syndrome. *European Respiratory Journal* 2003; 21(2):253-259.
55. Ficker JH, Wiest GH, Wilpert J, Fuchs FS, Hahn EG. Evaluation of a portable recording device (Somnocheck) for use in patients with suspected obstructive sleep apnoea. *Respiration* 2001; 68(3):307-312.
56. Fietze I, Glos M, Rottig J, Witt C. Automated analysis of data is inferior to visual analysis of ambulatory sleep apnea monitoring. *Respiration* 2002; 69(3):235-241.
57. Zucconi M, Ferini-Strambi L, Castronovo V, Oldani A, Smirne S. An unattended device for sleep-related breathing disorders: validation study in suspected obstructive sleep apnoea syndrome. *European Respiratory Journal* 1996; 9(6):1251-1256.
58. Vazquez JC, Tsai WH, Flemons WW, et al. Automated analysis of digital oximetry in the diagnosis of obstructive sleep apnoea. *Thorax* 2000; 55(4):302-307.
59. Pittman SD, Ayas NT, MacDonald MM, et al. Using a wrist-worn device based on peripheral arterial tonometry to diagnose obstructive sleep apnea: in-laboratory and ambulatory validation. *Sleep* 2004; 27(5):923-933.
60. Meoli AL, Casey KR, Clark RW, et al. Hypopnea in sleep-disordered breathing in adults. *Sleep* 2001; 24(4):469-470.
61. Hutter DA, Holland BK, Ashtyani H. Occult sleep apnea: the dilemma of negative polysomnography in symptomatic patients. *Sleep Med* 2004; 5(5):501-506.
62. Le Bon O., Hoffmann G, Tecco J, et al. Mild to moderate sleep respiratory events: one negative night may not be enough. *Chest* 2000; 118(2):353-359.
63. Quan SF, Griswold ME, Iber C, et al. Short-term variability of respiration and sleep during unattended nonlaboratory polysomnography--the Sleep Heart Health Study. *Sleep* 2002; 25(8):843-849.
64. Lettieri CJ, Eliasson AH, Andrada T, Khramtsov A, Kristo DA. Does zolpidem enhance the yield of polysomnography? *J Clin Sleep Med* 2005; 1(2):129-131.
65. Kingshott RN, Engleman HM, Deary IJ, Douglas NJ. Does arousal frequency predict daytime function? *Eur Respir J* 1998; 12(6):1264-1270.
66. Bennett LS, Langford BA, Stradling JR, Davies RJ. Sleep fragmentation indices as predictors of daytime sleepiness and nCPAP response in obstructive sleep apnea. *Am J Respir Crit Care Med* 1998; 158(3):778-786.
67. d'Ortho MP, Grillier-Lanoir V, Levy P, et al. Constant vs. automatic continuous positive airway pressure therapy: home evaluation. *Chest* 2000; 118(4):1010-1017.
68. Hermida RC, Zamarron C, Ayala DE, Calvo C. Effect of continuous positive airway pressure on ambulatory blood pressure in patients with obstructive sleep apnoea. *Blood Pressure Monitoring* 2004; 9(4):193-202.
69. Lloberes P, Marti S, Sampol G, et al. Predictive factors of quality-of-life

- improvement and continuous positive airway pressure use in patients with sleep apnea-hypopnea syndrome: study at 1 year. *Chest* 2004; 126(4):1241-1247.
70. Marrone O, Salvaggio A, Bonsignore MR, Insalaco G, Bonsignore G. Blood pressure responsiveness to obstructive events during sleep after chronic CPAP. *European Respiratory Journal* 2003; 21(3):509-514.
 71. Nosedá A, Kempnaers C, Kerkhofs M, Houben JJ, Linkowski P. Sleep apnea after 1 year domiciliary nasal-continuous positive airway pressure and attempted weight reduction. Potential for weaning from continuous positive airway pressure. *Chest* 1996; 109(1):138-143.
 72. Nosedá A, Jann E, Hoffmann G, Linkowski P, Kerkhofs M. Compliance with nasal continuous positive airway pressure assessed with a pressure monitor: pattern of use and influence of sleep habits. *Respiratory Medicine* 2000; 94(1):76-81.
 73. Whitelaw WA, Brant RF, Flemons WW. Clinical usefulness of home oximetry compared with polysomnography for assessment of sleep apnea.[see comment]. *American Journal of Respiratory & Critical Care Medicine* 2005; 171(2):188-193.
 74. Mulgrew AT, Fox N, Ayas NT, Ryan CF. Diagnosis and initial management of obstructive sleep apnea without polysomnography: a randomized validation study. *Annals of Internal Medicine* 2007; 146(3):157-166.
 75. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. *JAMA* 1999; 282(8):771-778.
 76. Ayappa I, Norman RG, Suryadevara M, Rapoport DM. Comparison of limited monitoring using a nasal-cannula flow signal to full polysomnography in sleep-disordered breathing. *Sleep* 2004; 27(6):1171-1179.
 77. Sokal RR, Rohlf EF. *Biometry*. New York: Freeman, 1981.
 78. Linnet K. Estimation of the linear relationship between the measurements of two methods with proportional errors. *Stat Med* 1990; 9(12):1463-1473.
 79. Bablok W, Passing H, Bender R, Schneider B. A general regression procedure for method transformation. Application of linear regression procedures for method comparison studies in clinical chemistry, Part III. *J Clin Chem Clin Biochem* 1988; 26(11):783-790.
 80. Ayas NT, Pittman S, MacDonald M, White DP. Assessment of a wrist-worn device in the detection of obstructive sleep apnea. *Sleep Medicine* 2003; 4(5):435-442.
 81. Baltzan MA, Verschelden P, Al-Jahdali H, Olha AE, Kimoff RJ. Accuracy of oximetry with thermistor (OxiFlow) for diagnosis of obstructive sleep apnea and hypopnea. *Sleep* 2000; 23(1):61-69.
 82. Gugger M, Mathis J, Bassetti C. Accuracy of an intelligent CPAP machine with in-built diagnostic abilities in detecting apnoeas: a comparison with polysomnography. *Thorax* 1995; 50(11):1199-1201.
 83. Zamarron C, Romero PV, Rodriguez JR, Gude F. Oximetry spectral analysis in the diagnosis of obstructive sleep apnoea. *Clinical Science* 1999; 97(4):467-473.
 84. 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(5):1567-1576.

85. Gagnadoux F, Pelletier-Fleury N, Philippe C, Rakotonanahary D, Fleury B. Home unattended vs hospital telemonitored polysomnography in suspected obstructive sleep apnea syndrome: a randomized crossover trial. *Chest* 2002; 121(3):753-758.
86. 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(3):536-540.
87. Mykytyn IJ, Sajkov D, Neill AM, McEvoy RD. Portable computerized polysomnography in attended and unattended settings. *Chest* 1999; 115(1):114-122.
88. Orr WC, Eiken T, Pegram V, Jones R, Rundell OH. A laboratory validation study of a portable system for remote recording of sleep-related respiratory disorders. *Chest* 1994; 105(1):160-162.
89. Portier F, Portmann A, Czernichow P, et al. Evaluation of home versus laboratory polysomnography in the diagnosis of sleep apnea syndrome. *American Journal of Respiratory & Critical Care Medicine* 2000; 162(3:Pt 1):t-8.
90. Ancoli-Israel S, Mason W, Coy TV, et al. Evaluation of sleep disordered breathing with unattended recording: the Nightwatch System. *Journal of Medical Engineering & Technology* 1997; 21(1):10-14.
91. Ballester E, Solans M, Vila X, et al. Evaluation of a portable respiratory recording device for detecting apnoeas and hypopnoeas in subjects from a general population. *European Respiratory Journal* 2000; 16(1):123-127.
92. Emsellem HA, Corson WA, Rappaport BA, et al. Verification of sleep apnea using a portable sleep apnea screening device. *Southern Medical Journal* 1990; 83(7):748-752.
93. Lloberes P, Montserrat JM, Ascaso A, et al. Comparison of partially attended night time respiratory recordings and full polysomnography in patients with suspected sleep apnoea/hypopnoea syndrome. *Thorax* 1996; 51(10):1043-1047.
94. Man GC, Kang BV. Validation of a portable sleep apnea monitoring device. *Chest* 1995; 108(2):388-393.
95. Marrone O, Salvaggio A, Insalaco G, Bonsignore MR, Bonsignore G. Evaluation of the POLYMESAM system in the diagnosis of obstructive sleep apnea syndrome. *Monaldi Archives for Chest Disease* 2001; 56(6):486-490.
96. Parra O, Garcia-Escasans N, Montserrat JM, et al. Should patients with sleep apnoea/hypopnoea syndrome be diagnosed and managed on the basis of home sleep studies?[see comment]. *European Respiratory Journal* 1997; 10(8):1720-1724.
97. Quintana-Gallego E, Villa-Gil M, Carmona-Bernal C, et al. Home respiratory polygraphy for diagnosis of sleep-disordered breathing in heart failure. *European Respiratory Journal* 2004; 24(3):443-448.
98. Su S, Baroody FM, Kohrman M, Suskind D. A comparison of polysomnography and a portable home sleep study in the diagnosis of obstructive sleep apnea syndrome. *Otolaryngology - Head & Neck Surgery* 2004; 131(6):844-850.
99. Verse T, Pirsig W, Junge-Hulsing B, Kroker B. Validation of the POLY-MESAM seven-channel ambulatory recording unit. *Chest* 2000; 117(6):1613-1618.
100. Whittle AT, Finch SP, Mortimore IL, Mackay TW, Douglas NJ. Use of home sleep studies for diagnosis of the sleep apnoea/hypopnoea syndrome. *Thorax* 1997; 52(12):1068-1073.

101. 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. *Otolaryngology - Head & Neck Surgery* 2006; 134(2):204-209.
102. Claman D, Murr A, Trotter K. Clinical validation of the Bedbug in detection of obstructive sleep apnea. *Otolaryngol Head Neck Surg* 2001; 125(3):227-230.
103. Adachi H, Mikami A, Kumano-go T, et al. Clinical significance of pulse rate rise during sleep as a screening marker for the assessment of sleep fragmentation in sleep-disordered breathing.[see comment]. *Sleep Medicine* 2003; 4(6):537-542.
104. Alvarez D, Hornero R, Abasolo D, del CF, Zamarron C. Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection. *Physiological Measurement* 2006; 27(4):399-412.
105. Bar A, Pillar G, Dvir I, et al. Evaluation of a portable device based on peripheral arterial tone for unattended home sleep studies.[see comment]. *Chest* 2003; 123(3):695-703.
106. Bonsignore G, Marrone O, Macaluso C, et al. Validation of oximetry as a screening test for obstructive sleep apnoea syndrome. *European Respiratory Journal - Supplement* 1990; 11:542s-544s.
107. Bradley PA, Mortimore IL, Douglas NJ. Comparison of polysomnography with ResCare Autoset in the diagnosis of the sleep apnoea/hypopnoea syndrome. *Thorax* 1995; 50(11):1201-1203.
108. Chiner E, Signes-Costa J, Arriero JM, et al. Nocturnal oximetry for the diagnosis of the sleep apnoea hypopnoea syndrome: a method to reduce the number of polysomnographies?[see comment]. *Thorax* 1999; 54(11):968-971.
109. Cooper BG, Veale D, Griffiths CJ, Gibson GJ. Value of nocturnal oxygen saturation as a screening test for sleep apnoea. *Thorax* 1991; 46(8):586-588.
110. Douglas NJ, Thomas S, Jan MA. Clinical value of polysomnography.[see comment]. *Lancet* 1992; 339(8789):347-350.
111. Golpe R, Jimenez A, Carpizo R. Home sleep studies in the assessment of sleep apnea/hypopnea syndrome. *Chest* 2002; 122(4):1156-1161.
112. Gurubhagavatula I, Maislin G, Nkwuo JE, Pack AI. Occupational screening for obstructive sleep apnea in commercial drivers. *American Journal of Respiratory & Critical Care Medicine* 2004; 170(4):371-376.
113. Gyulay S, Gould D, Sawyer B, et al. Evaluation of a microprocessor-based portable home monitoring system to measure breathing during sleep. *Sleep* 1987; 10(2):130-142.
114. Issa FG, Morrison D, Hadjuk E, et al. Digital monitoring of sleep-disordered breathing using snoring sound and arterial oxygen saturation. *American Review of Respiratory Disease* 1993; 148(4:Pt 1):t-9.
115. Levy P, Pepin JL, schaux-Blanc C, Paramelle B, Brambilla C. Accuracy of oximetry for detection of respiratory disturbances in sleep apnea syndrome. *Chest* 1996; 109(2):395-399.
116. 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. *Annals of Otology, Rhinology & Laryngology* 2006; 115(11):802-809.
117. Overland B, Bruskeland G, Akre H, Skatvedt O. Evaluation of a portable recording device (Reggie) with actimeter and nasopharyngeal/esophagus catheter incorporated. *Respiration* 2005; 72(6):600-605.

118. Pang KP, Dillard TA, Blanchard AR, et al. A comparison of polysomnography and the SleepStrip in the diagnosis of OSA. *Otolaryngology - Head & Neck Surgery* 2006; 135(2):265-268.
119. Penzel T, Kesper K, Pinnow I, Becker HF, Vogelmeier C. Peripheral arterial tonometry, oximetry and actigraphy for ambulatory recording of sleep apnea. *Physiological Measurement* 2004; 25(4):1025-1036.
120. Rauscher H, Popp W, Zwick H. Quantification of sleep disordered breathing by computerized analysis of oximetry, heart rate and snoring. *European Respiratory Journal* 1991; 4(6):655-659.
121. Rauscher H, Popp W, Zwick H. Model for investigating snorers with suspected sleep apnoea. *Thorax* 1993; 48(3):275-279.
122. Ryan PJ, Hilton MF, Boldy DA, et al. Validation of British Thoracic Society guidelines for the diagnosis of the sleep apnoea/hypopnoea syndrome: can polysomnography be avoided? *Thorax* 1995; 50(9):972-975.
123. Schafer H, Ewig S, Hasper E, Luderitz B. Predictive diagnostic value of clinical assessment and nonlaboratory monitoring system recordings in patients with symptoms suggestive of obstructive sleep apnea syndrome. *Respiration* 1997; 64(3):194-199.
124. Series F, Marc I, Cormier Y, La FJ. Utility of nocturnal home oximetry for case finding in patients with suspected sleep apnea hypopnea syndrome.[see comment]. *Annals of Internal Medicine* 1993; 119(6):449-453.
125. Stoohs R, Guilleminault C. MESAM 4: an ambulatory device for the detection of patients at risk for obstructive sleep apnea syndrome (OSAS). *Chest* 1992; 101(5):1221-1227.
126. Van Surell C., Lemaigre D, Leroy M, et al. Evaluation of an ambulatory device, CID 102, in the diagnosis of obstructive sleep apnoea syndrome. *European Respiratory Journal* 1995; 8(5):795-800.
127. 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(4):2166-2175.
128. Williams AJ, Yu G, Santiago S, Stein M. Screening for sleep apnea using pulse oximetry and a clinical score. *Chest* 1991; 100(3):631-635.
129. Bagnato MC, Nery LE, Moura SM, Bittencourt LR, Tufik S. Comparison of AutoSet and polysomnography for the detection of apnea-hypopnea events. *Brazilian Journal of Medical & Biological Research* 2000; 33(5):515-519.
130. Gugger M. Comparison of ResMed AutoSet (version 3.03) with polysomnography in the diagnosis of the sleep apnoea/hypopnoea syndrome. *European Respiratory Journal* 1997; 10(3):587-591.
131. Kiely JL, Delahunty C, Matthews S, McNicholas WT. Comparison of a limited computerized diagnostic system (ResCare Autoset) with polysomnography in the diagnosis of obstructive sleep apnoea syndrome. *European Respiratory Journal* 1996; 9(11):2360-2364.
132. Mayer P, Meurice JC, Philip-Joet F, et al. Simultaneous laboratory-based comparison of ResMed Autoset with polysomnography in the diagnosis of sleep apnoea/hypopnoea syndrome. *European Respiratory Journal* 1998; 12(4):770-775.
133. Rees K, Wraith PK, Berthon-Jones M, Douglas NJ. Detection of apnoeas, hypopnoeas and arousals by the AutoSet in the sleep apnoea/hypopnoea

- syndrome. *European Respiratory Journal* 1998; 12(4):764-769.
134. Shochat T, Hadas N, Kerkhofs M, et al. The SleepStrip: an apnoea screener for the early detection of sleep apnoea syndrome. *European Respiratory Journal* 2002; 19(1):121-126.
135. Kushida CA. Clinical presentation, diagnosis, and quality of life issues in restless legs syndrome. *Am J Med* 2007; 120(1 Suppl 1):S4-S12.
136. Allen RP, Picchietti D, Hening WA, et al. Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. *Sleep Med* 2003; 4(2):101-119.
137. Ohayon MM, Roth T. Prevalence of restless legs syndrome and periodic limb movement disorder in the general population. *J Psychosom Res* 2002; 53(1):547-554.
138. Mehra R, Strohl KP. Incidence of serious adverse events during nocturnal polysomnography.[see comment]. *Sleep* 2004; 27(7):1379-1383.
139. Bearpark H, Elliott L, Grunstein R, et al. Snoring and sleep apnea. A population study in Australian men. *American Journal of Respiratory & Critical Care Medicine* 1995; 151(5):1459-1465.
140. Flemons WW, Whitelaw WA, Brant R, Remmers JE. Likelihood ratios for a sleep apnea clinical prediction rule. *American Journal of Respiratory & Critical Care Medicine* 1994; 150(5:Pt 1):t-85.
141. Fry JM, DiPhillipo MA, Curran K, Goldberg R, Baran AS. Full polysomnography in the home. *Sleep* 1998; 21(6):635-642.
142. Hedner J, Pillar G, Pittman SD, et al. A novel adaptive wrist actigraphy algorithm for sleep-wake assessment in sleep apnea patients. *Sleep* 2004; 27(8):1560-1566.
143. Ip MS, Lam B, Tang LC, et al. A community study of sleep-disordered breathing in middle-aged Chinese women in Hong Kong: prevalence and gender differences. *Chest* 2004; 125(1):127-134.
144. Margel D, Cohen M, Livne PM, Pillar G. Severe, but not mild, obstructive sleep apnea syndrome is associated with erectile dysfunction. *Urology* 2004; 63(3):545-549.
145. Masa JF, Rubio M, Findley LJ. Habitually sleepy drivers have a high frequency of automobile crashes associated with respiratory disorders during sleep. *American Journal of Respiratory & Critical Care Medicine* 2000; 162(4:Pt 1):t-12.
146. Middelkoop HA, Knuistingh NA, van Hilten JJ, Ruwhof CW, Kamphuisen HA. Wrist actigraphic assessment of sleep in 116 community based subjects suspected of obstructive sleep apnoea syndrome. *Thorax* 1995; 50(3):284-289.
147. Neven AK, Middelkoop HA, Kemp B, Kamphuisen HA, Springer MP. The prevalence of clinically significant sleep apnoea syndrome in The Netherlands. *Thorax* 1998; 53(8):638-642.
148. Series F, Marc I. Nasal pressure recording in the diagnosis of sleep apnoea hypopnoea syndrome.[see comment]. *Thorax* 1999; 54(6):506-510.