Evidence Report

Obstructive Sleep Apnea and Commercial Motor Vehicle Driver Safety: Updated Review

Presented to

Federal Motor Carrier Safety Administration

11/30/2011

Prepared by:

MANILA Consulting Group, Inc.
1420 Beverly Road, Suite 220
McLean, VA 22101

Evidence reports are sent to the Federal Motor Carrier Safety Administration's (FMCSA) Medical Review Board (MRB) and Medical Expert Panels (MEP). The MRB and MEP make recommendations on medical topics of concern to FMCSA.

FMCSA will consider all MRB and MEP recommendations, however, all proposed changes to current standards and guidance (guidelines) will be subject to public-notice-and-comment and relevant rulemaking processes.

Authorship

Jessica R. Williams, PhD, MPH, RN	MANILA Consulting Group
Arit Amana, BA	MANILA Consulting Group
Stephen J. Tregear, DPhil	MANILA Consulting Group

Table of Contents

Executive Summary

Purpose of the Evidence Report

Driving is a complicated psychomotor performance that depends on fine coordination between the sensory and motor systems. Many health conditions exist which have the potential to impair perception, cognition (including alertness, attitude to risk, and recall) and/or motor function and, as a result, can make driving less safe.

Obstructive sleep apnea (OSA) is a relatively common disorder affecting approximately 12 million individuals in the United States, with approximately 4% of men and 2% of women in the U.S. suffering from symptomatic sleep apnea. [1-5] OSA is a disorder characterized by a reduction or cessation of breathing during sleep coupled with symptoms such as daytime sleepiness (i.e., OSA syndrome). [1, 2] Given this, OSA may culminate in unpredictable and sudden incapacitation (e.g., falling asleep at the wheel), thus contributing to the potential for crash, injury, and death.

In 2007, MANILA Consulting Group conducted a systematic review of the literature under the direction of the Department of Transportation's Federal Motor Carrier Safety Administration in order to synthesize the evidence related to OSA and crash risk, as well as the effectiveness of diagnostic tests and treatment options for OSA.

Since completion of this evidence report, a considerable amount of research has been conducted related to methods for the diagnosis of OSA. Much of this research has been conducted in response to a push by some to identify options for the diagnosis of OSA which could be used as an alternative to polysomnography (PSG). The purpose of this evidence report is to synthesize the research that has been conducted since the last review related to diagnostic alternatives to PSG for the identification of OSA. Specifically, this report focuses on two key questions:

Key Question# 1: Are there screening/diagnostic algorithms available that will enable examiners to identify those individuals at higher risk for moderate-to-severe OSA, thereby referring these individuals for confirmation by PSG?

Key Question #2: Are portable monitoring devices comparable to in-laboratory, technician-attended polysomnography (PSG) in the identification of individuals with OSA?

Identification of Evidence Bases

Separate evidence bases for each of the key questions addressed by this evidence report were identified using a process consisting of a comprehensive search of the literature, examination of abstracts of identified studies in order to determine which articles would be retrieved, and the selection of the actual articles that would be included in each evidence base.

A total of seven electronic databases (Medline, PubMed (pre Medline), EMBASE, PSYCH Info, CINAHL, TRIS, the Cochrane library) were searched (through April 30th, 2007). In addition, we examined the

reference lists of all obtained articles with the aim of identifying relevant articles not identified by our electronic searches. Hand searches of the "gray literature" were also performed. Admission of an article into an evidence base was determined by formal retrieval and inclusion criteria that were determined *a priori.*

Grading the Strength of Evidence

Our assessment of the quality of the evidence took into account not only the quality of the individual studies that comprise the evidence base for each key question; we also considered the interplay between the quality, quantity, robustness, and consistency of the overall body of evidence.

Analytic Methods

The set of analytic techniques used in this evidence report was extensive. Random- and fixed-effects meta-analyses were used to pool data from different studies. Differences in the findings of studies (heterogeneity) were identified using the Q-statistic and I^2. Sensitivity analyses, aimed at testing the robustness of our findings, included the use of cumulative fixed- and random-effects meta-analysis. The presence of publication bias was tested for using the "trim and fill" method.

Evidence-Based Conclusions

Key Question #1: Are there screening/diagnostic algorithms available that will enable examiners to identify those individuals at higher risk for moderate-to-severe OSA, thereby referring these individuals for confirmation by PSG?

Twelve articles met the inclusion criteria for Key Question #1. All 12 included studies measured the diagnostic performance of an algorithm/model developed to predict the presence and/or severity of OSA. The findings of our synthesis of these 12 studies are summarized below:

No recommendations can be made in support of any one algorithm as an appropriate screening tool to aid in OSA diagnoses. The algorithms investigated in this report (and any future algorithms developed) need to be tested among CMV drivers, in order to better determine their suitability in screening for moderate-to-severe OSA among this population.

There were several methodological issues encountered with the studies in our evidence base. First off, all algorithms investigated in this report were developed among non-realistic study populations, i.e. populations that did not mirror/approximate the CMV driver population in the United States. The study populations used to develop these algorithms were carefully-chosen, typically from among individuals presenting to sleep study centers with suspected disordered breathing and/or OSA. Males were generally underrepresented in the included studies, while rates of hypertension were overrepresented. In addition, the prevalence of moderate-to-severe OSA among all but two of the included studies ranged from 22.0% to 69.2% - significantly higher than the 10.5% of 400 commercial truck drivers found to have moderate-to-severe OSA by the American Transportation Research Institute (ATRI).

All but one of the algorithms in the included studies used at least one subjective variable. These variables (see were measured either by the self-report of the study participant or their bed partner. As stated earlier, self-reported data can be unreliable: respondent may lie or exaggerate their responses in order to please or appear more socially acceptable to the investigator. The resulting bias introduced by this limitation may skew the diagnostic performance of the algorithms under study.

Key Question #2: Are portable monitoring devices comparable to in-laboratory, technician-attended polysomnography (PSG) in the identification of individuals with OSA?

Fourteen articles provided evidence to inform the conclusions drawn from this updated systematic review examining the performance of PM devices in the diagnosis of OSA compared to the current gold standard, PSG. The findings of our analyses of these 14 studies are summarized below:

The findings of this updated systematic review support our previous findings that a number of portable sleep monitoring systems, though not as accurate as the current reference standard (PSG) do offer an alternative method by which the severity of PSA may be assessed in a large number of individuals at a relatively low cost.

Nine systematic reviews examining the performance of portable monitors in diagnosing OSA compared to PSG found that portable monitors performed reasonably well compared to PSG though none were as accurate (i.e., no PM device has a sensitivity and specificity of 100%). These reviews found that the majority of PM devices could differentiate individuals with OSA from those without and could differentiate individuals with severe OSA from those with mild-to-moderate OSA. Evidence was strongest for Level 3 PM devices for which more research has been conducted. Evidence does indicate that Level 2 and 4 devices show some utility, more research is needed to confirm these findings. Other findings from the systematic reviews indicate that manual scoring of PM devices provide results more consistent with PSG than automated scoring of PM devices; PM devices tend to result in more data loss than PSG although newer devices with built-in alert systems may help reduce these errors, and; PM tends to be associated with higher cost savings over PSG even when accounting for higher rates of data loss.

RCTs examining differences in clinical outcomes after CPAP treatment based OSA diagnosis with PM versus PSG, also support the utility of PM devices in the diagnosis of OSA. A variety of clinical outcomes were assessed across the four studies including AHI, sleepiness, quality of life, and functional and physical health. Very few differences were found between PM and PSG groups on any of these outcomes.

Three RCTs provided information in a manner that allowed us to conduct a meta-analysis. Specifically, we conducted a fixed effects meta-analysis to determine whether ESS scores after CPAP treatment differed between PM and PSG groups. The summary standardized difference in means was 0.129 (95% CI: -0.067, 0.335; p=0.325), suggesting a trend toward slightly better scores among the PSG group, although this difference was not statistically significant.

Preface

Background

Driving is a complicated psychomotor performance that depends on fine coordination between the sensory and motor systems. Many health conditions exist which have the potential to impair perception, cognition (including alertness, attitude to risk, and recall) and/or motor function and, as a result, can make driving less safe.

Obstructive sleep apnea (OSA) is a relatively common disorder affecting approximately 12 million individuals in the United States, with approximately 4% of men and 2% of women in the U.S. suffering from symptomatic sleep apnea. [1-5] OSA is a disorder characterized by a reduction or cessation of breathing during sleep coupled with symptoms such as daytime sleepiness (i.e., OSA syndrome). [1, 2] Given this, OSA may culminate in unpredictable and sudden incapacitation (e.g., falling asleep at the wheel), thus contributing to the potential for crash, injury, and death.

In 2007, MANILA Consulting Group conducted a systematic review of the literature under the direction of the Department of Transportation's Federal Motor Carrier Safety Administration in order to synthesize the evidence related to OSA and crash risk, as well as the effectiveness of diagnostic tests and treatment options for OSA.

Since completion of this evidence report, a considerable amount of research has been conducted related to methods for the diagnosis of OSA. Much of this research has been conducted in response to a push by some to identify options for the diagnosis of OSA which could be used as an alternative to polysomnography (PSG). The purpose of this evidence report is to synthesize the research that has been conducted since the last review related to diagnostic alternatives to PSG for the identification of OSA. Specifically, this report focuses on two key questions:

Key Question 1: Are there screening/diagnostic algorithms available that will enable examiners to identify those individuals at higher risk for moderate-to-severe OSA, thereby referring these individuals for confirmation by PSG?

Key Question 2: Are portable monitoring devices comparable to in-laboratory, technician-attended polysomnography (PSG) in the identification of individuals with OSA?

Organization of Report

This evidence report contains three major sections: 1) *Methods*, 2) *Evidence Synthesis,* and 3) *Conclusions*. These major sections are supplemented by extensive use of appendices.

In the *Methods* section, we detail how we identified and analyzed information for this report. The section covers the key questions addressed, details of literature searching, criteria for including studies in our analyses, evaluation of study quality, assessment of the strength of the evidence base for each

question, and methods for abstracting and synthesis of clinical study results. The *Evidence Synthesis* section of this report is organized by key question. For each question, we provide a brief background related to the key question being addressed. Readers are referred to the previous systematic review for more in depth background information related to OSA. This section also report on the quality and quantity of the studies that provided relevant evidence. We then summarize available data extracted from included studies either qualitatively or, when the data permit, qualitatively and quantitatively (using meta-analysis). Each section in the Evidence Synthesis closes with our conclusions that are based on our assessment of the available evidence. This evidence report ends with a *Conclusions* section that briefly summarizes the answers to each of the questions addressed.

Methods

The *Methods* section provides a synopsis of how we identified and analyzed information for this report. The section briefly covers the key questions addressed, literature searches performed, the criteria used including studies, evaluation of study quality, assessment of the strength of the evidence base for each key question, and the methods used for abstracting and analyzing available data. Specific details of literature searches, study quality assessment, statistical approaches used, etc. are documented in appendices.

Key Questions

This evidence report addresses two key questions. Each of these key questions was developed by FMCSA such that the answers to these questions provided information that would be useful in updating their current medical examination guidelines. The key questions addressed in this evidence report are as follows:

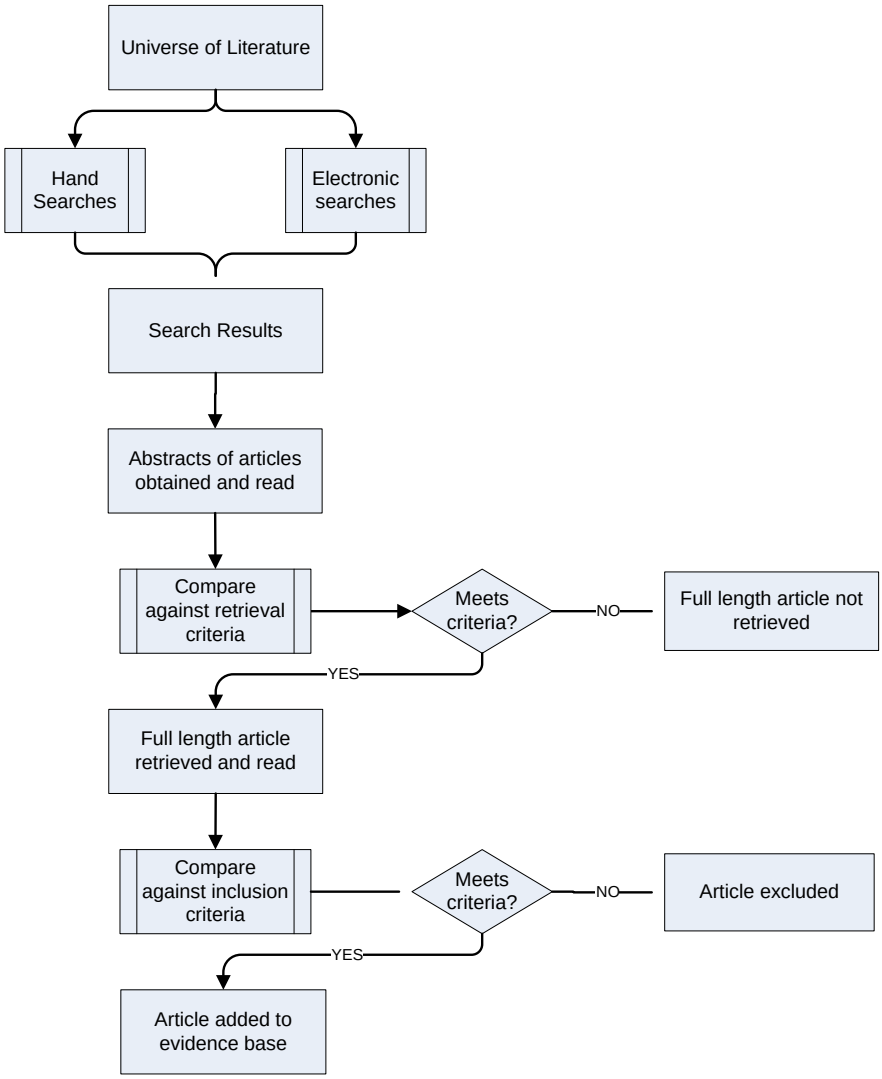
Key Question 1: Are there screening/diagnostic algorithms available that will enable examiners to identify those individuals at higher risk for moderate-to-severe OSA, thereby referring these individuals for confirmation by PSG?

Key Question 2: Are portable monitoring devices comparable to in-laboratory, technician-attended polysomnography (PSG) in the identification of individuals with OSA?

Identification of Evidence Bases

The individual evidence bases for each of the key questions addressed in this evidence report were identified using the multistage process captured by the algorithm presented in Figure 1. The first stage of this process consists of a comprehensive search of the literature. The second stage of the process consists of the examination of abstracts of identified studies in order to determine which articles will be retrieved. The final stage of the process consists of the selection of the actual articles that will be included in the evidence base.

Figure 1. Evidence Base Identification Algorithm



Searches

One characteristic of a good evidence report is a systematic and comprehensive search for information. Such searches distinguish systematic reviews from traditional literature reviews which use a less rigorous approach to identifying and obtaining literature thereby allowing a reviewer to include only articles that agree with a particular perspective and to ignore articles that do not. Our approach precludes this potential reviewer bias because we obtain and include articles according to explicitly determined *a priori* criteria. Full details of the search strategies used in this report are presented in Appendix A.

Electronic Searches

We performed comprehensive searches of the electronic databases listed in Table 1.

Table 1. Electronic Databases Searched

Name of database	Date limits	Platform/provider
CINAHL (Cumulative Index to Nursing and Allied Health Literature)	2003 through April 30, 2007	OVID
The Cochrane Database of Systematic Reviews (Cochrane Reviews)	2003 through 2007 Issue 2	www.thecochranelibrary.com
Database of Abstracts of Reviews of Effects (DARE)	2003 through 2007 Issue 2	www.thecochranelibrary.com
The Cochrane Central Register of Controlled Trials (CENTRAL)	2003 through 2007 Issue 2	www.thecochranelibrary.com
The Cochrane Database of Methodology Reviews (Methodology Reviews)	2003 through 2007 Issue 2	www.thecochranelibrary.com
ECRI Institute Library Catalog	2003 through 2007	ECRI Institute
Embase (Excerpta Medica)	2003 through April 30, 2007	OVID
Health Technology Assessment Database (HTA)	2003 through 2007 Issue 2	www.thecochranelibrary.com
Medline	2003 through April 30, 2007	OVID
National Guideline Clearinghouse (NGC)	2003 through April 30, 2007	www.ngc.gov
NHS Economic Evaluation Database (NHS EED)	2003 through 2007 Issue 2	www.thecochranelibrary.com
PsycINFO	2003 through April 30, 2007	OVID
PubMed (Pre Medline)	Premedline[sb} Searched March 30, 2007	www.pubmed.gov
TRIS Online (Transportation Research Information Service Database)	Searched April 30, 2007	http://ntlsearch.bts.gov/tris/index.do

Manual Searches

Non-journal publications and conference proceedings from professional organizations, private agencies, and government agencies were screened. In addition, we examined the reference lists of all obtained articles with the aim of identifying relevant reports not identified by our electronic searches. In order to retrieve additional relevant information, we also performed hand searches of the "gray literature." Gray literature consists of reports, studies, articles, and monographs produced by federal and local government agencies, private organizations, educational facilities, consulting firms, and corporations. The latter documents do not appear in the peer-reviewed journal literature.

Retrieval Criteria

Retrieval criteria were used to determine whether a full-length version of an article identified by our searches should be ordered. Decisions pertaining to whether a full-length article should be retrieved are

usually based on a review of available abstracts. For this project, retrieval criteria were determined *a priori* in conjunction with FMCSA. The retrieval criteria are presented in Appendix B.

If an article did not meet the retrieval criteria for this evidence report, the full-length version of the article was not obtained. If it was unclear whether a potentially relevant article met our retrieval criteria (e.g., no abstract was available for evaluation), the full-length version of that article was be obtained.

Inclusion and Exclusion Criteria

Each retrieved article was read in full by a research analyst who determined whether that article met a set of predetermined, question specific, inclusion criteria. As was the case for the retrieval criteria, the inclusion criteria for this evidence report were determined *a priori* in conjunction with FMCSA. These inclusion and exclusion criteria are presented in Appendix C.

If on reading an article it was found not to meet the question specific inclusion criteria listed in Appendix C, the article was excluded from the analysis. Each excluded article, along with the reason(s) for its exclusion, are presented in Appendix D.

Evaluation of Quality and Strength of Evidence

Rather than focus on the quality of the individual studies that comprise an evidence base, our approach to assessing the quality of evidence focused on the overall *body* of the available evidence that was used to draw an evidence-based conclusion. [6] Using this approach, which is described briefly in Appendix E, we took into account not only the quality of the individual studies that comprise the evidence base for each key question, we will also consider the interplay between the quality, quantity, robustness, and consistency of the overall body of evidence.

Our approach to assessing the strength of the body of evidence makes a clear distinction between a qualitative conclusion (e.g., "Individuals with OSA are at increased risk for a motor vehicle crash") and a quantitative conclusion (e.g., "When compared to individuals who do not have OSA, the risk ratio for a motor vehicle crash among individuals with the disorder is 1.37; 95% CI: 1.03–1.74; P <0.005."). As shown in Table 2, we assigned a separate strength of evidence rating to each of type of conclusion. Evidence underpinning a qualitative conclusion was rated according to its strength, and evidence underpinning quantitative conclusions was rated according to the stability of the effect size estimate that was calculated.

Table 2. Strength of Evidence Ratings for Qualitative and Quantitative Conclusions

Strength of Evidence	Interpretation
Qualitative Conclusion	
Strong	Evidence supporting the qualitative conclusion is convincing. It is highly unlikely that new evidence will lead to a change in this conclusion.
Moderate	Evidence supporting the qualitative conclusion is somewhat convincing. There is a small chance that new evidence will overturn or strengthen our conclusion. ECRI recommends regular monitoring of the relevant literature for moderate-strength conclusions.
Minimally	Although some evidence exists to support the qualitative conclusion, this evidence is tentative and perishable. There is a reasonable

acceptable	chance that new evidence will either overturn or strengthen our conclusions. ECRI recommends frequent monitoring of the relevant literature.
Unacceptable	Although some evidence exists, the evidence is insufficient to warrant drawing an evidence-based conclusion. ECRI recommends frequent monitoring of the relevant literature.
Quantitative Conclusion (Stability of Effect Size Estimate)	
High	The estimate of treatment effect in the conclusion is stable. It is highly unlikely that the magnitude of this estimate will change substantially as a result of the publication of new evidence.
Moderate	The estimate of treatment effect the conclusion is somewhat stable. There is a small chance that the magnitude of this estimate will change substantially as a result of the publication of new evidence. ECRI recommends regular monitoring of the relevant literature.
Low	The estimate of treatment effect included in the conclusion is likely to be unstable. There is a reasonable chance that the magnitude of this estimate will change substantially as a result of the publication of new evidence. ECRI recommends frequent monitoring of the relevant literature.
Unstable	Estimates of the treatment effect are too unstable to allow a quantitative conclusion to be drawn at this time. ECRI recommends frequent monitoring of the relevant literature.

The definitions presented in the table above are intuitive. Qualitative conclusions that are supported by strong evidence are less likely to be overturned by the publication of new data than conclusions supported by weak evidence. Likewise, quantitative effect size estimates that deemed to be stable are more unlikely to change significantly with the publication of new data than are unstable effect size estimates.

Statistical Methods

The set of analytic techniques used in this report was extensive. In summary, random- and fixed-effects meta-analyses were used to pool appropriate data from different studies. [7-15] Important differences in the findings of different studies (heterogeneity) were identified using the Q-statistic and I^2. [12, 16-21] Whenever appropriate, heterogeneity was explored using meta-regression techniques. [22-24] Sensitivity analyses were used to test the robustness of all findings. The presence of publication bias was tested for using the "trim and fill" method. [25] All meta-analyses in this Evidence Report were performed using Comprehensive Meta-Analysis software. [26]

We calculated several different estimates of effect. The choice of effect size estimate depended on the purpose of the studies we assessed, their design, and whether reported outcome data were continuous or dichotomous. Between-group differences in outcome measured using continuous data were analyzed in their original metric (if all included studies reported on the same outcome using the same metric) or the data were standardized into a common metric known as the standardized mean difference (SMD). Dichotomous data were analyzed using the rate ratio (RR) or the odds ratio (OR). Time-to-event data were analyzed using the hazard ratio (RH). The formulae for these effect sizes and their variance are presented in Table 3. If means and standard deviations were not available for continuous data, every effort was made to determine an estimate of treatment effect from reported statistics (e.g., t-values, f-values) or from p-values using methods described in detail elsewhere. [27]

Table 3. Effect Size Estimates Used in Evidence Report and their Variance

Effect size	Formula (Effect size)	Formula (Variance)
WMD	$\mu_{TG} - \mu_{CG}$	$\left(\sqrt{\frac{(n_{TG}-1)(s_{TG})^2+(n_{CG}-1)(s_{CG})^2}{n_{TG}+n_{CG}-2}}\right)\left(\frac{1}{n_{TG}}+\frac{1}{n_{cg}}\right)$
SMD	$\frac{\mu_{TG}-\mu_{CG}}{\left(\sqrt{\frac{(n_{TG}-1)(s_{TG})^2+(n_{CG}-1)(s_{CG})^2}{n_{TG}+n_{CG}-2}}\right)}$	$\frac{n_{TG}+n_{CG}}{n_{TG}n_{CG}}+\frac{SMD^2}{2(n_{TG}+n_{CG})}$
Where: μ_{TG} = mean (treatment group); μ_{CG} = mean (control group); S_{TG} = standard deviation (treatment group); S_{CG} = standard deviation (control group); n_{TG} = enrollees (treatment group); n_{CG} = enrollees (control group)		
Event Rate	$a/a+b$	$\ln\left[\frac{1}{a}+\frac{1}{a+b}\right]$
Where: a = number of individuals in cohort experiencing an event; b = number of individuals in cohort who did not experience an event		
RR (incidence)	$\left(\frac{a_{OSAs}}{pt_{OSA}}\right) \Big/ \left(\frac{b_{control}}{pt_{control}}\right)$	$\ln\left[\frac{1}{a_{OSA}}+\frac{1}{b_{control}}\right]$
Where: a = number of individuals with OSA who crashed; pt_{OSA} = rate denominator (OSA grp); b = number of individuals without OSA who crashed; $pt_{control}$ = rate denominator (control grp)		
OR	$\left(\frac{a}{b}\right) \Big/ \left(\frac{c}{d}\right) = \left(\frac{ad}{bc}\right)$	$\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}$
RR	$\left(\frac{a}{a+c}\right) \Big/ \left(\frac{b}{b+d}\right)$	$\frac{1}{a}+\frac{1}{a+c}+\frac{1}{b}+\frac{1}{b+d}$
Where: a = number of individuals with OSA who crashed; b = number of individuals without OSA who crashed; c = number of individuals with OSA who did not crash; d = number of individuals without OSA who did not crash.		
HR	$\frac{O_{pi}/E_{pi}}{O_{ci}/E_{ci}}$	$\exp\left(\ln\left[\frac{1}{E_{pi}}+\frac{1}{E_{ci}}\right]\right)$

Effect size	Formula (Effect size)	Formula (Variance)
Where O_{pi} = observed number of events in treatment group; O_{ci} = observed number of events in control group; E_{pi} = logrank expected number of events in treatment group; E_{ci} = logrank expected number of events in control group		

OSA = obstructive sleep apnea; HR = hazard ratio; RR = rate ratio; OR = Odds Ratio; RR = rate ratio; SMD = standardized mean difference; WMD = weighted mean difference

Evidence Synthesis

KEY QUESTION #1: Are there screening/diagnostic algorithms available that will enable examiners to identify those individuals at higher risk for moderate-to-severe OSA, thereby referring these individuals for confirmation by PSG?

BACKGROUND

Moderate-to-Severe OSA among Drivers

Moderate-to-severe OSA is defined by an apnea-hypopnea index (AHI) of greater than-or-equal-to 15 [28]. OSA is up to about 10 times more prevalent among commercial drivers, compared to the general population, [17] and the American Transportation Research Institute found a 10.5% prevalence of moderate-to-severe OSA among commercial truck drivers.[29] Individuals with OSA have been shown to be between 21% and 389% more likely to experience motor vehicle crash, compared to individuals without the condition [21]. Continuous positive airway pressure (CPAP) is a treatment widely recommended for patients with OSA, especially when other more conservative treatments have failed [17]. CPAP has been shown to reduce crash risk among drivers with moderate-to-severe OSA [17]. In addition, CPAP also relieves excessive daytime sleepiness associated with OSA within one day of treatment [17].

Diagnosis of Moderate-to-Severe OSA

Polysomnography (PSG) is currently considered the gold standard in identifying individuals with moderate-to-severe OSA, who would be appropriate candidates for CPAP treatment [30]. PSG (also referred to as "sleep study") measures various physiological factors in a sleeping subject. PSG may be attended or unattended; during attended PSG, a technician observes the sleeping subject and monitors the recording equipment. A typical PSG session includes [31]:

- Electroencephalogram (EEG)
- Electro-oculogram (EOG)
- Electromyogram (EMG)
- Oral and nasal airflow measurement
- Chest and abdominal movement measurement
- Audio recording of snoring loudness
- Oximetry (measurement of blood oxygen saturation)
- Video monitoring of subject

After a PSG session concludes, the collected data are analyzed by a board-certified sleep specialist. The number of apneas, hypopneas, leg movements, oxygen desaturations, and sleep levels are formally reported, and a diagnosis is made [31].

The use of PSG in OSA diagnosis presents significant problems related to access and cost [32]. For example, the cost of a PSG session has been found to be over $4000.00 [33]. The associated costs of OSA diagnosis by PSG may prove especially problematic for individuals without adequate health insurance.

Criteria for Referring Commercial Motor Vehicle (CMV) Drivers to PSG

According to Section 49 CFR 391.41 (b) (5) of the Federal Motor Carrier Safety Regulations (FMCSA), CMV drivers are required to undergo medical qualification examinations every two years [32]. Hartenbaum et al. (2006) suggested several updates to the FMCSA's 1991 guidelines on the evaluation of drivers with suspected OSA. According to their recommendations, CMV drivers should undergo an in-service evaluation if they fall into any of the following categories:

- A sleep history suggestive of OSA (snoring, excessive daytime sleepiness, witnessed apneas)
- Two or more of the following: Body Mass Index (BMI) > 35 kg/m2; neck circumference > 17 inches in men and 16 inches in women; hypertension (new, uncontrolled or unable to be controlled with less than two medications)
- An Epworth Sleepiness Scale score > 10
- A previously diagnosed sleep disorder; compliance claimed, but no recent medical visits/compliance data available for immediate review (must be reviewed within 3-month period); if found not to be compliant, should be removed from service (includes surgical treatment)
- An AHI of > 5 but < 30 in a prior PSG, together with an Epworth sleepiness scale score < 11, no motor vehicle accidents and no hypertension requiring two or more agents to control

For CMV drivers meeting any one of the following categories, Hartenbaum et al. (2006) recommend an immediate out-of-service evaluation:

- Observed unexplained excessive daytime sleepiness (sleeping in examination or waiting room) or confessed excessive sleepiness.
- Motor vehicle accident (run off road, at fault, rear-end collision) likely related to sleep disturbance unless evaluated for sleep disorder in the interim.
- Motor vehicle accident (run off road, at fault, rear-end collision) likely related to sleep disturbance unless evaluated for sleep disorder in the interim.
- Previously diagnosed sleep disorder (1) noncompliant (continuous positive airway pressure treatment not tolerated); (2) no recent follow-up (within recommended time frame); (3) any surgical approach with no objective follow-up.
- An AHI > 30.

As part of a Medical Expert Panel (MEP), Ancoli-Israel et al. (2008) recommended to the FMCSA that commercial drivers meeting ANY of the following criteria should undergo an evaluation to confirm the diagnosis of OSA:

- Categorized as high risk for OSA according to the Berlin Questionnaire
- BMI ≥ 33 kg/m2

- Judged to be at high risk for OSA based on a clinical evaluation

Implementing the above recommendations for determining PSG referrals for commercial drivers may prove problematic because they could result in large numbers of drivers being referred to PSG [34]. This would increase the demand on already over-stretched PSG centers, leading to significant delays in OSA diagnoses, and a corresponding delay in commercial drivers being cleared to work.

Prioritizing Commercial Drivers for PSG by the Use of Screening Algorithms

Prioritizing commercial drivers with a higher risk of moderate-to-severe OSA could be helpful in minimizing the costs of OSA diagnosis by PSG. One expects that a higher proportion of these priority individuals would be diagnosed with OSA [35], thereby reducing the frequency of "needless" PSG, i.e. PSG procedures performed on individuals without OSA.

There are several factors that may indicate an increased risk of having OSA. OSA diagnosis by PSG may be worthwhile for individuals with one or more of these risk factors. To this end, several investigators (using these risk factors as variables) have attempted to develop algorithms (or models) able to predict the presence and/or severity of OSA. This section investigates a number of these algorithms, and assesses their predictive power in identifying individuals with moderate-to-severe OSA.

The value of algorithms lies in their ability to correctly identify those people that will and will NOT be diagnosed as having OSA. As an example, consider Algorithms A and B. Of a group of 100 individuals, both algorithms predict that 50 people have a higher chance of having OSA. The 50 high-risk people identified by Algorithm A undergo further testing; 27 of them are found to have OSA. After further testing, 34 of the 50 high-risk people identified by Algorithm B are diagnosed with OSA. Hence, Algorithm B is more useful than Algorithm A in correctly identifying individuals with a higher chance of having OSA, i.e. Algorithm B has a higher ***sensitivity*** than Algorithm A. The 50 people identified as low-risk by Algorithm A undergo testing; 4 of them are found to have OSA. After also undergoing further testing, 11 of the 50 low-risk individuals identified by Algorithm B are diagnosed with OSA. Algorithm A is more useful than Algorithm B in correctly identifying people with a lower chance of having OSA, i.e. Algorithm A has a higher ***specificity*** than Algorithm B. Algorithms with higher sensitivity values result in less resource "waste" during further testing for OSA presence/severity, while algorithms with higher specificity values help in avoiding the negative consequences of individuals with OSA going undiagnosed.

Algorithms with higher sensitivity AND specificity values are considered more superior than those with lower sensitivities and specificities.

According to the literature, the most salient risk factors for OSA are presented below. The presence of more than one risk factor in an individual may or may not indicate a higher risk for OSA.

- Age [32, 36-44]
- Alcohol Use [45]
- Body-mass index (BMI) [1, 5-9, 32, 36-40, 42-44, 46-58]

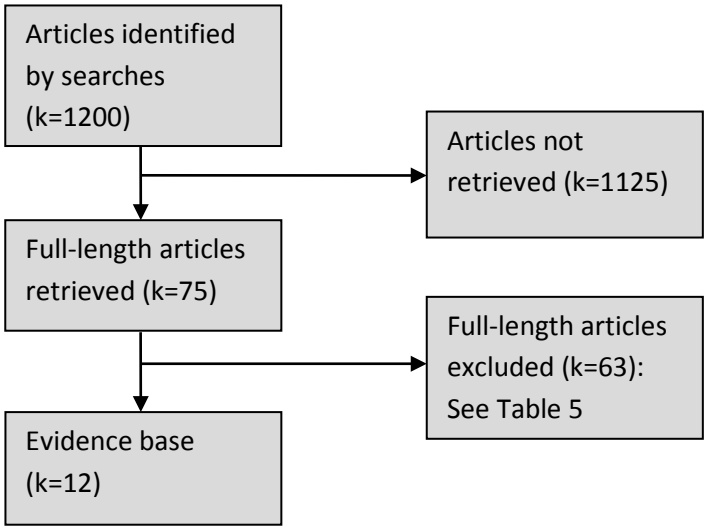
- Coronary Disease [32, 38]
- Daytime Sleepiness [38]
- Diabetes [32, 38]
- Epworth Sleepiness Scale (ESS) Score [40, 41, 59-62]
- Head & Neck Anatomical Abnormalities [4, 5, 37, 42-44, 52, 63-65]
- Hypertension [32, 38]
- Male sex [32, 36-44, 66]
- Mallampati Grade [5, 40]
- Neck circumference [5, 9, 36, 39, 41, 45, 50, 52, 55-58]
- Observed Apneas [36, 38, 39, 67]
- Snoring [36, 38, 40, 41, 67, 68]
- Waist Circumference [9, 56-58, 69, 70]

Identification of Evidence Base

In order to address Key Question #1, we searched for studies that evaluated the ability of any algorithm or model to identify individuals with moderate-to-severe OSA, as verified by a reference standard (PSG).

The evidence base was developed via, 1) a comprehensive search of the literature (through August 2011), 2) the examination of the abstracts of identified articles to determine which articles would be retrieved, and 3) the selection of the actual articles that would form the evidence base (Figure 2). To supplement the electronic searches, we also examined the references/bibliographies of the included studies, recent narrative reviews, and selected grey literature sources. The retrieval of an article, and its subsequent admission into the evidence base, was determined by specific retrieval and inclusion criteria, as listed in Appendix B and Appendix C, respectively.

Figure 2. Development of Evidence Base for Key Question #1



Our searches identified a total of 1200 articles that appeared to be relevant to Key Question #1. Following application of the retrieval criteria, 75 full-length articles were retrieved and read in full. Of these 75 articles, 12 articles describing 12 unique studies met the inclusion criteria for Key Question #1. Appendix D lists the 63 articles that were retrieved but then excluded and provides the primary reason for their exclusion.

Table 4 lists the 12 articles that met the inclusion criteria for Key Question #1. All 12 included studies measured the diagnostic performance of an algorithm/model developed to predict the presence and/or severity of OSA.

Table 4. Evidence Base for Key Question #1

Reference	Year	Study Location	Country
Chen et al. [71]	2010	Taiwan	China
Crocker et al. [72]	1990	Newcastle NSW	Australia
Dixon et al. [73]	2003	Melbourne	Australia
Khoo et al. [74]	2010	Singapore	Singapore
Morris et al. [75]	2008	New York	USA
Pillar et al. [54]	1994	Haifa	Israel
Pradhan et al. [76]	1996	Massachusetts	USA
Rauscher et al. [77]	1993	Vienna	Austria

Rowley et al. [78]	2000	Michigan	USA
Sharma et al. [79]	2004	New Delhi	India
Sharma et al. [80]	2006	New Delhi	India
Viner et al. [81]	1991	Ontario	Canada

Evidence Base

This subsection provides a brief description of the key attributes of the 12 studies that comprise the evidence base for Key Question #1. Here we discuss applicable information on the quality of the included studies and the generalizability of each study's findings to CMV drivers. The key attributes of the algorithms presented in each included study are presented in Table 5.

Table 5. Key Attributes of Algorithms Used in Included Studies

Reference	# Variables in model	List of Variables	Equation/Model	Threshold(s)	Sensitivity; Specificity; PPV; NPV and/or AUC	Prevalence of moderate-to-severe OSA (AHI ≥ 15)
Chen et al. [71]	5 (in OSA model)	▪ Sex ▪ Age ▪ BMI ▪ ESS ▪ Sleep Outcomes Survey (SOS) score	Probability of OSA = ex / 1 + ex where x = (-5.935 + 1.096*sex + 0.064*age + 0.264*BMI + 0.039*ESS – 0.062*SOS) sex = 1 if male, 0 if female	RDI ≥ 5	At SOS ≤ 55 and ESS ≥ 9: Sensitivity = 60.3% Specificity = 72.9% PPV = 93.43% NPV = 22.29%	67.0%
Crocker et al. [72]	4	▪ Stopped breathing ▪ Hypertension ▪ BMI ▪ Age	1/(1 + e-(-13.9 + 0.06a + 2.98b + 0.23c + 1.35d)) where a = age in years, b = 1 when apnea is reported and 0 when not, c = BMI in kg/m2, d = 1 when hypertension is present and 0 when not	AHI > 15	At cut-off of 0.15: Sensitivity = 85.0% Specificity = 61.0%	46.7%
Dixon et al. [73]	6	▪ BMI ≥ 45 ▪ Age ▪ Observed sleep apnea ▪ HBA1c ≥ 6% ▪ Fasting plasma insulin ≥ 28 μmol/L ▪ Male sex	BASH'IM Score (range = 0 to 6)	AHI ≥ 15	At BASH'IM Score cutoff of ≥ 3: Sensitivity = 89.0% Specificity = 81.0% PPV = NR NPV = NR AUC: 0.91	44.4%
				AHI ≥ 30	At BASH'IM Score cutoff of ≥ 3: Sensitivity = 96.0% Specificity = 71.0% PPV = NR NPV = NR AUC: 0.92	

Reference	# Variables in model	List of Variables	Equation/Model	Threshold(s)	Sensitivity; Specificity; PPV; NPV and/or AUC	Prevalence of moderate-to-severe OSA (AHI ≥ 15)
Khoo et al. [74]	4	▪ Neck circumference ≥ 40cm ▪ Male sex ▪ Frequent awakening with unrefreshing sleep ▪ Age ≥ 50 years	Probability of having OSA = 1/(1 + exp (-z)) where z = -2.74 + (0.95*Age≥50) + (1.31*male) + (1.78*neck circumference ≥ 40cm) + (1.29*frequent awakening with unrefreshing sleep) 0 = absence of variable; 1 = presence of variable	AHI ≥ 20	At cut-off of 0.6: Sensitivity = 77.9% Specificity = 72.5% PPV = 84.5% NPV = 63.0% AUC = 0.792	65.8% (with AHI ≥ 20)
Morris et al. [75]	2	▪ Snoring Severity Scale (SSS) score ▪ BMI	For SSS ≥ 4 or BMI ≥ 26, there is increased risk of moderate-to-high OSA	RDI ≥ 15	For SSS ≥ 4 or BMI ≥ 26: Sensitivity = 97.4% Specificity = 40.0% PPV = 82.3% NPV = 84.2% AUC = 0.82 (SSS score) AUC = 0.71 (BMI)	69.2%
Pillar et al. [54]	4	▪ Self-report of apneas (SRA) ▪ Neck circumference index (NCI) ▪ Tendency to fall asleep unintentionally (TFAU) ▪ Age	Predicted Apnea Index (pAI) = -131.5 + 11.67*SRA + 1.02*NCI + 5.78*TFAU + 0.04*Age SRA and TFAU given in units: 1 = never; 2 = seldom; 3 = frequently; 4 = always NCI = (1000*neck circumference [cm])/(310 + 55*height[m])	AHI > 10	Sensitivity = 92.2% Specificity = 18.2% PPV = 76.6% NPV = NR	NR
				AHI > 20	Sensitivity = 81.6% Specificity = 48.6% PPV = NR NPV = NR	
Pradhan et al. [76]	Clinical data model: 4	▪ Sex ▪ BMI ▪ Frequency of loud snoring as reported on the PSQI	Probability of apnea = ek/1 + ek where k = (-7.92 – 1.35a + 0.14b + 0.49c + 0.05d); a = 0 if male, 1 if	RDI > 10	At probability cut-off of 0.16: Sensitivity = 100% Specificity = 16%	22% (with AHI > 30)

Reference	# Variables in model	List of Variables	Equation/Model	Threshold(s)	Sensitivity; Specificity; PPV; NPV and/or AUC	Prevalence of moderate-to-severe OSA (AHI ≥ 15)
		▪ Age	female; b = BMI; c = frequency of snoring, 0 to 3 times/week; age in years			
Rauscher et al. [77]	4	▪ Reported episodes of apnea ▪ >20% overweight ▪ Male sex ▪ Falling asleep while reading	Probability of OSA = ek/(1 + ek) where k= (-7.263 – 2.046a + 0.0487b + 1.121c + 1.663d); a = 0 if male and 1 if female, b = (weight[kg] x 100)/(height [cm] - 100), c = 1 if falling asleep while reading is reported and 0 if not, d = 1 if episodes of apnea are reported and 0 if not.	AHI ≥ 10	Sensitivity = 94% Specificity = 45% NPV (probability below 0.31 for AHI ≤ 10) = 91%	NR
				AHI ≥ 20	Sensitivity = 95% Specificity = 41% NPV (probability below 0.31 for AHI ≤ 20) = 94%	
Sharma et al. [79]	3	▪ Gender ▪ Waist/Hip Ratio (WHR) ▪ Neck Circumference (NC)	Score = 1.378*gender + 0.064*WHR + 0.21*NC where gender = 0 if female and 1 if male; WHR = percentage of normal WHR taken as 0.85; NC = actual neck circumference of subject	AHI ≥ 15	At score cutoff of > 16.62: Sensitivity = 90.4% Specificity = 69.8% PPV = 71.2% NPV = 89.8%	44.9%
Sharma et al. [80]	4	▪ Gender ▪ Snoring Index ▪ Choking Index ▪ BMI	Score = 1.61*gender + 1.01*snoring index + 2.09*choking index + 0.1*BMI where gender =o if female and 1 if male	AHI ≥ 15	At cutoff value of 4.3: Sensitivity = 91.3% Specificity = 68.5% PPV = 70.5% NPV = 92.3% AUC = 0.896	45%

Reference	# Variables in model	List of Variables	Equation/Model	Threshold(s)	Sensitivity; Specificity; PPV; NPV and/or AUC	Prevalence of moderate-to-severe OSA (AHI ≥ 15)
STUDIES ON EFFECTIVENESS OF OTHER MODELS						
Rowley et al. [78] (models tested: Crocker et al. [72]; Viner et al. [81])	Crocker Model: 4	▪ Age ▪ Witnessed apneas ▪ BMI ▪ Hypertension	Probability of OSA = 1/(1 + e-(-13.9 + 0.06a + 2.98b + 0.23c + 1.35d)) where a = age; b = 1 if witnessed apneas present, 0 if absent; c = BMI; d = 1 if hypertensive, 0 if not	AHI ≥ 10	At cutoff of 0.15: Sensitivity = 84% Specificity = 39% PPV = 73% AUC = 0.669	49.0% (with AHI ≥ 20)
				AHI ≥ 20	At cutoff of 0.95: Sensitivity = 33% Specificity = 90% PPV = 76% AUC = 0.700	
	Viner Model: 4	▪ Sex ▪ Age ▪ BMI ▪ Snoring	Probability of OSA = ex/(1 + ex) where x = -10.5132 + 0.9164*sex + 0.0470*age + 0.1869*BMI + 1.932*snoring Where sex = 1 for male, 0 for female; snoring = 1 for present, 0 for absent	AHI ≥ 10	At cutoff of 0.20: Sensitivity = 96% Specificity = 13% PPV = 69% AUC = 0.695	
				AHI ≥ 20	At cutoff of 0.95: Sensitivity = 34% Specificity = 87% PPV = 72% AUC = 0.722	

Abbreviations: AHI = apnea hypo-apnea index; AUC = area under receiver-operator curve; BASH'IM = acronym representing BMI, Age, Observed sleep apnea, HbA1c, fasting plasma insulin, and male gender; BMI = body mass index; ESS = Epworth Sleepiness Scale; Hb1Ac = Hemoglobin A1c; NPV = negative predictive value; OSA = obstructive sleep apnea; PPV = positive predictive value; PSQI = Pittsburg Sleep Quality Index

Quality of Included Studies

In the following section, we summarize our quality assessment of the included articles, using the QUADAS tool (see Table 6).[82] We present our assessment results in tabular format, and follow with a discussion of certain issues brought to light by the quality assessment. We do not provide summary quality scores for each study; summary scores can mask the reality that quality-score components may vary in importance from study to study.[83] Blinding, for example, impacts effect sizes more in a pain study than in a mortality study, but summary scoring would weight blinding equally in these two study types.

Table 6. Quality of the Studies that address Key Question #1

	Was the spectrum of patients representative of the patients who will receive the test in practice?	Were selection criteria clearly described?	Is the reference standard likely to correctly classify the target condition?	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	Did patients receive the same reference standard regardless of the index test result?	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	Was the execution of the index test described in sufficient detail to permit replication of the test?	Was the execution of the reference standard described in sufficient detail to permit its replication?	Were the index test results interpreted without knowledge of the results of the reference standard?	Were the reference standard results interpreted without knowledge of the results of the index test?	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	Were uninterruptable/intermediate test results reported?	Were withdrawals from the study explained?
Chen et al. [71]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NR	NR	Y	Y	N
Crocker et al. [72]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NR	NR	Y	Y	N
Dixon et al. [73]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NR	NR	Y	Y	NR
Khoo et al. [74]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NR	Y	Y	Y	Y
Morris et al. [75]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NR	Y	Y	N
Pillar et al. [54]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NR	NR	Y	Y	N
Pradhan et al. [76]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NR	NR	Y	Y	N
Rauscher et al. [77]	N	Y	Y	Y	Y	Y	Y	N	Y	NR	NR	Y	Y	N

	Was the spectrum of patients representative of the patients who will receive the test in practice?	Were selection criteria clearly described?	Is the reference standard likely to correctly classify the target condition?	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	Did patients receive the same reference standard regardless of the index test result?	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	Was the execution of the index test described in sufficient detail to permit replication of the test?	Was the execution of the reference standard described in sufficient detail to permit its replication?	Were the index test results interpreted without knowledge of the results of the reference standard?	Were the reference standard results interpreted without knowledge of the results of the index test?	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	Were uninterruptable/intermediate test results reported?	Were withdrawals from the study explained?
Rowley et al. [78]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NR	NR	Y	Y	Y
Sharma et al. [79]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NR	NR	Y	Y	Y
Sharma et al. [80]	Y	Y	Y	Y	Y	Y	Y	Y	Y	NR	Y	Y	Y	Y
Viner et al. [81]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	NR	Y	Y	N

Y = Yes; N = No; NR = Not Reported

Inferences from Quality Assessment of Included Studies

Spectrum of Study Participants

All but one of the included studies recruited study participants that reflected a range of OSA severity levels, from mild OSA (AHI ≥ 5, < 15) through to severe OSA (AHI ≥ 30). In the study by Rauscher et al. (1993), 95 of the 184 study participants had AHI values below 10, while the rest of the participants had AHI values above 20.[77] There were no study participants representing OSA severity as indicated by AHI values between 10 and 20, thus giving rise to spectrum bias and resulting in an over-estimation of the algorithm's performance.

Accuracy of Reference Standard

All the included studies utilized PSG as their reference standard in determining the diagnostic values of their OSA-prediction algorithms. The assumption is that PSG has 100% sensitivity and 100% specificity. However, the accuracy and performance of PSG has been disputed. Almost all of the signals recorded during PSG are uncalibrated, and the manufacture of PSG equipment components is not standardized.[84] PSG scoring is manual and based largely on pattern recognition of qualitative signals;[84] this may contribute to the night-to-night variability observed in respiratory abnormalities measured by PSG. [85, 86] In addition, studies have found poor correlation between any of the variables measured by PSG and patients' symptoms and/or treatment outcomes.[87-92] For example, patients with slightly-increased AHI values have presented with severe daytime sleepiness, while those with high AHI values have shown little-to-no symptoms.[93, 94]

Blinding

All the included studies had a situation where the technicians scoring algorithm results were NOT blinded to the subject's corresponding PSG result, or vice versa. Consequently, significant bias may have been introduced into these studies' algorithm/PSG scoring, thereby lowering one's confidence in the algorithm's diagnostic abilities.

Study Withdrawals

Only 4 of the 12 included studies disclosed the number of withdrawals from their study and/or explained the reasons for these dropouts.[74, 78-80] One study (Dixon et al. 2003) did not report withdrawals but did compare study participants (who were enrolled consecutively) to those who were not enrolled.[73] The rest of the included studies merely reported the final study population size, without references to possible withdrawals. If there were systematic differences between any unexplained withdrawals and retained subjects in these studies, bias may have been introduced, thereby lowering one's confidence in the diagnostic abilities of the algorithm being studied.

Studies' Description of Algorithms and PSG

One of the 12 included studies failed to describe their execution of either the index test (algorithm) or the reference standard used (PSG) in a manner that allowing replication.[77] In the study by Rauscher et al. (1993), the authors mention using a 36-item questionnaire to gather data on their algorithm's

variables, but they did not reproduce this questionnaire, or inform readers of where it could be obtained.

Generalizability of Evidence to Target Population

Important characteristics of the individuals included in the studies that address Key Question #1 are summarized in Table 7. As a whole, the generalizability of the individuals enrolled in the included studies to CMV drivers is unclear. Among 1,329 CMV drivers surveyed by Gurubhagavatula et al. (2004), the average age was 44.4 (± 11.2) years, 93.5% were male, and the average BMI was 28.4 (± 4.85) kg/m2.[34] The rate of obesity (≥ 30 kg/m2) among 103 CMV drivers was found to be 53.4%; [95] Smith & Phillips (2001) found that 69.6% of 595 CMV drivers had a BMI equal to or over 30 kg/m2 and 47.6% of these drivers had a BMI above 33 kg/m2.[96] Among the studies included in this report, the average age of participants ranged from 40.4 to 49.8 years – somewhat comparable to the population surveyed by Gurubhagavatula et al. (2004), while the proportion of males (ranging from 24.0% to 93.0%) was less comparable. The lowest BMI values reported in the included studies were also comparable to that of the Gurubhagavatula et al. (2004) study. Five of the included studies reported average BMI values above 30 kg/m2; indicating prevalent obesity among these study populations (not unlike the Wiegand (2009) and Smith & Phillips (2011) studies).[73, 75, 77-79]

Of the 1,329 drivers surveyed by Gurubhagavatula et al. (2004), 406 underwent PSG; the distribution of AHI values among these drivers was as follows: AHI < 5: 71.9%; ≤ 5 AHI < 15: 17.6%; ≤ 15 AHI < 30: 5.8%; AHI ≥ 30: 4.7%.[34] With the exception of one study,[72] the proportion of participants in the 13 included studies increased with higher AHI/RDI values (indicating increasing OSA severity) – a trend opposite to that in the Gurubhagavatula et al. (2004) study. One explanation for this is that these studies enrolled participants referred for the diagnosis/treatment of suspected sleep disorders. These kinds of individuals would present with more severe symptoms.

Five of the 12 included studies reported hypertension rates among their study participants.[72-74, 78, 80] The rates of hypertension reported in these five studies (ranging from 39.2% to 61.5%) are unlike that reported by Smith & Phillips (2011), who found an 8.7% hypertension rate among 595 commercial motor drivers.[96]

Five of the 12 included studies reported the mean neck circumference of their participants, [73, 74, 78-80] ranging from 37.3cm to 43.4cm (see Table 7). These values are a little below the average neck circumference of 45.5cm reported by Talmage et al. (2008) among 134 CMV drivers. [97] Parks et al. (2009) reported an average neck circumference of 41.6cm among 394 CMV Drivers, [35] which is more similar to the range of neck circumference values reported in the included studies.

Table 7. Individuals in Studies that address Key Question #1

Reference	Number of Participants (N)	Age (years)	BMI (kg/m2)	Males (%)	Neck Circumference (cm)	Hypertensive (%)	Distribution of AHI/RDI values (%)
Chen et al.	355	Mean = 44.7 (±	Mean =	87.9	Mean = NR	NR	RDI < 5: 13.5

Reference	Number of Participants (N)	Age (years)	BMI (kg/m2)	Males (%)	Neck Circumference (cm)	Hypertensive (%)	Distribution of AHI/RDI values (%)
[71]		11.3) Range = 18-80	27.4 ± 4.1 Range = NR		Range = NR		≤ 5 RDI < 15: 19.4 ≤ 15 RDI < 30: 14.6 RDI ≥ 30: 52.4
Crocker et al. [72]	100	Mean = 51.5 (SEM = 1.4)	Mean = 29.3 (SEM = 0.5)	83.0	NR	42.0	AHI ≤ 15: 73.0 AHI > 15: 27.0
Dixon et al. [73]	99	Mean = 40.4 (± 10.4) Range = 20-60	Mean = 47.6 (± 8.9) Range = 32.0 – 76.0	24.0	Mean = 43.4 (± 4.2) Range = NR	54.0	AHI < 5: 28.3 ≤ 5 AHI < 15: 35.4 ≤ 15 AHI < 30: 13.1 AHI ≥ 30: 23.2
Khoo et al. [74]	117	Mean = 47.5 (± NR) Range = NR	Mean = 28.5 (± NR) Range = NR	78.6	Mean = 40.5 (± NR) Range = NR	61.5	AHI < 20: 34.2 AHI ≥ 20: 65.8
Morris et al. [75]	211	Mean = 47.5 (± 16.3) Range = NR	Mean = 30.2 (± 5.9) Range = NR	69.7	Mean = NR Range = NR	NR	RDI < 15: 30.8 RDI ≥ 15: 69.2
Pillar et al. [54]	86	Mean = 47.0 (± 13) Range = NR	Mean = 29.1 (± 4.6) Range = NR	93.0	Mean = NR Range = NR	NR	Mean AHI = 30.5 (± 25.5)
Pradhan et al. [76]	150	Mean = 45.9 (± NR) Range = NR	Mean = NR Range = NR	71.3	Mean = NR Range = NR	NR	RDI < 10: 43.3 ≤ 10 RDI < 30: 34.7 ≤ 30 RDI < 50: 4.7 RDI ≥ 50: 17.3
Rauscher et al. [77]	184	Mean = 49.8 (± NR)	Mean = 30.0 (± NR)	75.0	Mean = NR Range = NR	NR	AHI < 10: 51.6 AHI > 20: 48.4

Reference	Number of Participants (N)	Age (years)	BMI (kg/m2)	Males (%)	Neck Circumference (cm)	Hypertensive (%)	Distribution of AHI/RDI values (%)
		Range = NR	Range = NR				
Rowley et al. [78]	370	Mean = 46.0 (± NR) Range = 36.0 – 56.8	Mean = 37.3 (± NR) Range = 29.6 - 49.4	52.0	Mean = 41.6 (± NR) Range = 36.0 – 48.0	51.4	AHI < 10: 33.0 ≤ 10 AHI < 20: 17.6 AHI ≥ 20: 49.4
Sharma et al. [79]	118	Mean = 45.9 (± NR) Range = NR	Mean = 34.0 (± NR) Range = NR	57.6	Mean = 39.6 (± NR) Range = NR	NR	AHI < 15: 55.1 AHI ≥ 15: 44.9
Sharma et al. [80]	102	Mean = 41.3 (± NR) Range = NR	Mean = 27.1 (± NR) Range = NR	68.6	Mean = 37.3 (± NR) Range = NR	39.2	AHI < 15: 54.9 AHI ≥ 15: 45.1
Viner et al. [81]	410	Mean = 45.7 (± NR) Range = NR	Mean = 28.4 (± NR) Range = NR	82.4	Mean = NR Range = NR	NR	< 10 AHI ≤ 30: 24.0 < 30 AHI ≤ 50: 8.0 AHI > 30: 10.0

Abbreviations: AHI = Apnea-Hypopnea Index; NR = Not Reported; RDI = Respiratory Disturbance Index

Performance Characteristics of Algorithms in Included Studies

We encountered great variability among the algorithms (both the algorithm equations and their variables) in the included studies. No two studies used the same algorithm or the same set of variables, preventing the direct comparison of any number of algorithms (see Appendix A). We decided to focus on three criteria that we believe matter most when screening for moderate-to-severe OSA in a clinical setting: the correct classification of individuals with moderate-to-severe OSA, the algorithm's clinical utility, and the algorithm's diagnostic performance.

The Correct Classification of Individuals with Moderate-to-Severe OSA

As stated previously, moderate-to-severe OSA is defined by an AHI ≥ 15 [28]. The algorithms in the 12 included studies used varying AHI thresholds (ranging from AHI ≥ 5 to AHI ≥ 30) to define the presence of OSA among their participants. An algorithm with an AHI threshold of ≥ 10 (for example) would identify individuals with moderate-to-severe OSA; however, the resulting group sent for confirmation by PSG would also contain those with mild OSA. Conversely, all individuals screened by an algorithm with

an AHI threshold of ≥ 20 will be classified as having moderate-to-severe OSA, but some individuals also having moderate-to-severe OSA will be missed.

One may reason that increasing the threshold at which a given algorithm has an acceptable diagnostic performance will only improve the algorithm's sensitivity and specificity. However, given two populations (diseased and non-diseased) in which the distribution of disease status is normal, increasing an algorithm's threshold will always (1) misclassify a greater number of diseased individuals as non-diseased, and (2) classify more non-diseased individuals correctly. The former effect shrinks the true positive value, resulting in a lower sensitivity, while the latter effect increases the true negative value, resulting in a higher specificity.

Of the 12 included studies, five studies used an AHI threshold of ≥ 15 to signify the presence of OSA among their participants: [72, 73, 75, 79, 80]

- Crocker at al. (1990)
- Dixon et al. (2003)
- Morris et al. (2008)
- Sharma et al. (2006)
- Sharma et al. (2004)

Clinical Utility

In this report, we adopt the term "clinical utility" to describe the ease of using the algorithms in the included studies. An algorithm is considered to have better clinical utility if its variables are (1) measurable without special equipment, training (beyond that of the typical clinician) or the need for offsite testing/interpretation, and (2) objectively measured (see Table 8). The need for special tools, training or offsite testing (by an external laboratory, for instance), could significantly impede the use of an algorithm in a clinical setting. The most serious issue with self-reported data concerns the validity and accuracy of participants' responses.[98] There's no alternative but to assume that participants' really behave/feel the way they say they do; however, participants may lie or exaggerate in order to present themselves in the best light possible, according to prevailing social norms/values. [98] The five articles identified in the preceding section (those that correctly classify screened individuals with moderate-to-severe OSA) are highlighted in Table 8.

Table 8. Clinical Utility of Algorithm Variables

	< - Variables representing better Clinical Utility - >													Variables representing Lower Clincal Utility	
	< - - - - - - Objectively-measured - - - - - - - -						< - - - - - - - - - - Subjectively-measured - - - - - - - - - - >								
Reference	Age	BMI	Sex	Hyper-tension	Neck circum-ference	Waist-Hip Ratio	Apneas	Choking	ESS	Falling Asleep	Frequent waking	SOS	Snoring	HbA1c	Insulin (plasma)
Chen et al. (2010)	✓	✓	✓						✓			✓			
Crocker et al. (1990)	✓	✓		✓			✓								
Dixon et al. (2003)	✓	✓	✓				✓							✓	✓
Khoo et al. (2010)	✓		✓		✓						✓				
Morris et al. (2008)		✓											✓		
Pillar et al. (1994)	✓				✓		✓			✓					
Pradhan et al. (1996)	✓	✓	✓										✓		
Rauscher et al. (1993)		✓	✓				✓			✓					
Sharma et al. (2004)			✓		✓	✓									
Sharma et al. (2006)		✓	✓					✓					✓		
Viner et al. (1991)	✓	✓	✓										✓		
TOTAL	7	8	8	1	3	1	4	1	1	2	1	1	4	1	1

Out of the five studies highlighted in Table 8, only the algorithm by Sharma et al. (2004) contained no subjective variables. [79] The algorithms by Crocker et al. (1990), Dixon et al. (2003) & Morris et al. (2008) had one subjective variable each (self-reported apneas or smoking). [72, 73, 75] The algorithm by Sharma et al. (2006) contained two subjective variables (self-reported smoking and snoring).[80] It is noteworthy that only the algorithm by Dixon et al. (2003) contained variables of lower clinical utility (*Hb1Ac* and *plasma insulin*).[73] These variables involve the analysis of blood samples, which may necessitate the use of external laboratories. Presently, these two variables are not typically measured as part of the Medical Examination required for all CMV drivers every two years. [32] This could change in the future, however, and warrant a reevaluation of *Hb1Ac* and *plasma insulin* as low clinical utility variables.

Diagnostic Performance

The sensitivities and specificities of the five algorithms discussed in preceding sections ranged from 85.0% to 97.4%, and from 40.0% to 81.0%, respectively (see Table 9 and Appendix A). As discussed in the Background section, algorithms with high sensitivities could reduce the number of needless PSG sessions conducted on individuals with suspected moderate-to-severe OSA. Conversely, algorithms with high specificities could reduce the number of individuals misdiagnosed as being without moderate-to-severe OSA, when in actuality, they have OSA.

Table 9. Diagnostic Performance of Selected Algorithms

Reference	Sensitivity	Specificity
Crocker et al. [72]	85.0%	61.0%
Dixon et al. [73]	89.0%	81.0%
Morris et al. [75]	97.4%	40.0%
Sharma et al. [79]	90.4%	69.8%
Sharma et al. [80]	91.3%	68.5%
Rowley et al. [78]	84.0% (for AHI ≥ 10); 33.0% (for AHI ≥ 20)	39.0% (for AHI ≥ 10); 90.0% (for AHI ≥ 20)

If used to screen 100 people, 30 of whom had PSG-confirmed moderate-to-severe OSA, the algorithm by Morris et al. (2008) would misdiagnose just one of these individuals – 4 people less than would be misdiagnosed by the Crocker et al. (1990) algorithm, and 2 people less than would be misdiagnosed by the Dixon et al. (2003), Sharma et al. (2004) & Sharma et al. (2006) algorithms.[72, 73, 75, 79, 80] However, using the Morris et al. (2008) algorithm would also result in 42 needless PSG sessions, compared to just 13 needless PSGs from use of the Dixon et al. (2003) algorithm.[73, 75] The Sharma et al. (2004), Sharma et al. (2006) & Crocker et al. (1990) algorithms would result in 21, 22 & 27 needless PSGs respectively.[72, 79, 80]

Rowley et al. (2000) studied the algorithm developed by Crocker et al. (1990), using a dissimilar study population (see Table 7) and an AHI threshold of ≥ 10 and ≥ 20 (instead of ≥ 15).[78] An AHI threshold of ≥ 10 would result in the classification of individuals with mild OSA as having moderate-to-severe OSA, while an AHI threshold of ≥ 20 would result in the exclusion of people with moderate-to-severe OSA. The sensitivity and specificity of the Rowley et al. (2000) algorithm at an AHI threshold ≥ 10 was 84.0% and 39.0% respectively; at an AHI threshold of ≥ 20, the algorithm had 33% sensitivity and 90% specificity.[78]

Continuing the screening scenario described above, the Rowley et al. (2000) algorithm (for an AHI ≥ 10) would misdiagnose the same number of individuals (5 people) as being free from moderate-to-severe OSA (when compared to the Crocker et al. (1990) algorithm), but result in 16 additional needless PSGs (43 total).[72, 78] For an AHI ≥ 20, the Rowley et al. (2000) algorithm would misdiagnose 20 individuals as being free from moderate-to-severe OSA, but result in just 7 needless PSGs, when compared to the Crocker et al. (1990) algorithm.[72, 78]

DISCUSSION

There were several methodological issues encountered with the studies in our evidence base. First off, all algorithms investigated in this report were developed among non-realistic study populations, i.e. populations that did not mirror/approximate the CMV driver population in the United States. The study populations used to develop these algorithms were carefully-chosen, typically from among individuals presenting to sleep study centers with suspected disordered breathing and/or OSA. Males were

generally underrepresented in the included studies, while rates of hypertension were overrepresented. In addition, the prevalence of moderate-to-severe OSA among all but two [54, 77] of the included studies ranged from 22.0% to 69.2% - significantly higher than the 10.5% of 400 commercial truck drivers found to have moderate-to-severe OSA by the American Transportation Research Institute (ATRI).[29]

All but one [79] of the algorithms in the included studies used at least one subjective variable. These variables (see Table 8) were measured either by the self-report of the study participant or their bed partner. As stated earlier, self-reported data can be unreliable: respondent may lie or exaggerate their responses in order to please or appear more socially acceptable to the investigator.[98] The resulting bias introduced by this limitation may skew the diagnostic performance of the algorithms under study.

In terms of quality, all the included studies had issues with blinding: either the algorithm scorer was not blinded to the results of participants' PSG, vice versa or both. This lack of blinding in the interpretation of algorithm/PSG results could have biased the diagnostic performances of the algorithms under study. In addition, 8 of the included studies did not account for possible study withdrawals, which also may have biased the algorithms' diagnostic performances.

CONCLUSION

Based on the issues raised in the Discussion section, we are unable to recommend any one algorithm as an appropriate screening tool to aid in OSA diagnoses. The algorithms investigated in this report (and any future algorithms developed) need to be tested among CMV drivers, in order to better determine their suitability in screening for moderate-to-severe OSA among this population.

Key Question 4: Are there screening/diagnostic tests available that will enable examiners to identify those individuals with OSA who are at an increased risk for a motor vehicle crash?

Background

The purpose of this section is to update the findings of our previous assessment as to whether there are screening/diagnostic tests available that will enable examiners to identify those individuals with OSA who are at an increased risk for a motor vehicle crash. Specifically, this update will focus on the performance of portable sleep monitoring devices compared to the current reference standard for diagnosing and determining the severity of OSA, in-laboratory, technician-attended polysomnography (PSG).

Polysomnography

Among other physiological parameters such as air flow, heart rate and rhythm, and respiratory effort, PSG assesses all of the known risk factors for motor vehicle crash. This has led to suggestions that all individuals who wish to be certified to drive a CMV and are suspected of, or diagnosed with, OSA, should undergo overnight PSG at a specialist sleep center. For example, the September 2006 recommendations regarding the evaluation for fitness-for-duty from the Joint Task Force of the American College of Chest Physicians, American College of Occupational Health and Environmental Medicine, and the National Sleep Foundation [99] state that all those wishing to drive a CMV and who are suspected of having sleep apnea should be assessed by a sleep physician and have any diagnosis confirmed by overnight. The recommendations define an individual who is suspected of having OSA as meeting one or more of the following criteria:

1. A sleep history suggestive of OSA (snoring, excessive daytime sleepiness, witnessed apneas)
2. Two or more of the following:
    a. BMI≥35 kg/m^2
    b. Neck circumference ≥17 inches in men or 16 inches in women
    c. ESS score ≥10
    d. Previous diagnosis of sleep apnea and no information on compliance with treatment

Coupled with these recommendations is a growing awareness among physicians and medical examiners of the danger that OSA poses to transportation safety. Together, these factors will increase the demand for access to sleep labs which will be difficult to satisfy in the face of an acknowledged shortage of testing facilities. This shortfall may lead to delays in diagnosis and treatment initiation. In addition to the deficit in sleep labs, the cost for a PSG is high, and may limit access to appropriate testing. [100-102] Consequently, alternative strategies to PSG that can detect and measure the severity of the known risk factors for a crash are actively being considered.

One such alternative to PSG is "split-night polysomnography". The initial diagnostic portion of the study, which is necessary to confirm the presence and severity of OSA, is followed on the same night by CPAP titration. The advantage of a split-night study is a presumed decrease in cost, because the two tests are administered in one night, rather than two. This alternative to the traditional PSG, while potentially faster and more cost effective than full PSG, does not overcome the problems of limited resources; the patient must still attend a sleep lab. Additional alternative testing modalities have been suggested, including clinical prediction models, portable sleep monitoring devices that can be used at home, and the use of various psychometric instruments primarily aimed at measuring sleepiness or attentiveness in the office.

Portable Sleep Monitoring Devices

Portable sleep monitoring (PM) is defined as a sleep study that is performed outside of the setting of a sleep laboratory. The term portable monitoring includes a wide range of devices that can be as complex as PSG (and measure all of the same parameters) or straightforward in that they assess only one parameter, such as oxygen saturation (oximetry).

The American Academy of Sleep Medicine (AASM) has defined four types (levels) of sleep testing based on the environment, technician attendance and number of parameters recorded (Table 10).[103] Portable sleep monitoring systems are classified as Level 2, 3, or 4 monitoring devices.

Table 10. AASM Sleep Monitor Categories

Category	Portability	Parameters Measured
Level 1	In-laboratory attended standard PSG. Measure both respiratory and sleep variables	Minimum 7 parameters: EEG, EOG, chin EMG EKG or heart rate, airflow, respiratory efforts, oxygen saturation
Level 2	Comprehensive Portable Full PSG performed in the home Measure both respiratory and sleep variables	Monitors the same channels as level 1 but not in a sleep lab Minimum 7 parameters: EEG, EOG, chin EMG EKG or heart rate, airflow, respiratory efforts, oxygen saturation
Level 3	Modified Portable Assessment of cardiorespiratory variables only	Minimum 4 parameters including: ventilation (2 channels of respiratory movements or respiratory movements and airflow), heart rate or EKG and oxygen saturation
Level 4	Portable Single or double parameter recordings	Minimum one parameter, usually oximetry alone or with one other channel such as airflow

A wide variety of Level 2 to Level 4 sleep monitoring systems are currently available in the United States. Most of these systems contain software which allows automated analysis and scoring of recorded signals.[104]

The potential advantages of home studies include convenience, improved access to testing, lower cost compared with in-laboratory studies, and the familiar sleeping environment afforded to individuals undergoing testing. In some cases, data transfer is made via modem to the laboratory analysis station, where the signal quality can then be assessed and equipment problems quickly addressed if required.[105]

While theoretically the costs of operating portable sleep monitoring devices are lower than laboratory based programs, many of the currently available devices require set up to take place in a laboratory, or require technical assistance in the home.[101] In the latter case, the costs associated with home monitoring are not much different to those associated with testing in a sleep lab.[106] When one takes into account the fact that portable equipment is more prone to damage and sleep studies are more likely to be inconclusive or fail (meaning that these failed studies will need to be repeated) the costs associated with sleep studies based on portable systems may ultimately exceed those associated with assessment in a sleep lab.

Historical Perspective on the Use of Unattended Portable Monitors in the Diagnosis of Obstructive Sleep Apnea in Adult Patients

Several organizations have attempted to develop practice parameters on the use of PSG and/or PM in OSA diagnosis and treatment. The first of such parameters was published in 1994, by the American Sleep Disorders Association.[107] In 1997, Chesson et al. published a paper on the indications for the use of PSG.[108] The Agency for Healthcare Research and Quality (AHRQ) also performed a meta-analysis of the literature on the diagnosis of OSA in 2000.[109] In 2003, a committee composed of representatives from AASM, the American College of Chest Physicians and the American Thoracic Society, published practice parameters based on this previous evidence review.[3] None of these publications supported the use of PM in OSA diagnosis, due to lack of sufficient evidence.

In 2005, the Centers for Medicare and Medicaid Services (CMS) released a statement that the current evidence was inadequate to conclude that "the use of unattended portable multi-channel sleep testing with a minimum of 4 or 7 monitored channels was reasonable and necessary in the diagnosis of OSA; therefore these tests remain uncovered."[110]

In 2006, the AASM released a position statement, recommending that physicians choosing to use PM do so within the context of the patient's comprehensive evaluation; that such devices only be utilized by board-certified sleep specialists, or at AASM-accredited sleep centers or laboratories; and that decisions regarding treatment be based on thorough evaluation of PM results and patient's symptoms.[111]

In 2007, CMS initiated a review of their previous statement at the request of the American Academy of Otolaryngology – Head and Neck Surgery [112]. The AASM, American College of Chest Physicians, and American Thoracic Society all argued that there was a lack of evidence on the efficacy of PM in the Medicare population and a lack of economic data in support of PM. AASM also reiterated its position that if PM is accepted as a diagnostic tool, it must be performed by certified sleep medicine specialists or by AASM-accredited facilities.

The AASM also charged the Portable Monitoring Task Force with reevaluating the evidence on PM as a suitable alternative to in-laboratory PSG. The Task Force performed a search to capture relevant articles published since its last systematic review in 2003,[3] and used an evidence review and consensus process to develop clinical guidelines for the use of PM in the diagnosis and management of OSA. Table 11 presents a summary of these guidelines.

Table 11. 2007 AASM Guidelines for the Use of Portable Monitors in the Diagnosis of OSA

1. Indications for Portable Monitoring

1.1. PM for the diagnosis of OSA should be performed only in conjunction with a comprehensive sleep evaluation. Clinical sleep evaluations using PM must be supervised by a practitioner with board certification in sleep medicine or an individual who fulfills the eligibility criteria for the sleep medicine certification examination. In the absence of a comprehensive sleep evaluation, there is no indication for the use of PM.

1.2. Provided that the recommendations of 1.1 have been satisfied, PM may be used as an alternative to Polysomnography (PSG) for the diagnosis of OSA in patients with a high pretest probability of moderate to severe OSA. PM should not be used in the patient groups described in 1.2.1, 1.2.2, and 1.2.3 (those with comorbidities, other sleep disorders, or for screening).

1.2.1. PM is not appropriate for the diagnosis of OSA in patients with significant comorbid medical conditions that may degrade the accuracy of PM, including, but not limited to, moderate to severe pulmonary disease, neuromuscular disease, or congestive heart failure.

1.2.2. PM is not appropriate for the diagnostic evaluation of OSA in patients suspected of having other sleep disorders, including central sleep apnea, periodic limb movement disorder (PLMD), insomnia, parasomnias, circadian rhythm disorders, or narcolepsy.

1.2.3. PM is not appropriate for general screening of asymptomatic populations.

1.3. PM may be indicated for the diagnosis of OSA in patients for whom in-laboratory PSG is not possible by virtue of immobility, safety, or critical illness.

1.4. PM may be indicated to monitor the response to non-CPAP treatments for obstructive sleep apnea, including oral appliances, upper airway surgery, and weight loss.

2. Technology for Portable Monitors

2.1. At a minimum, the PMs must record airflow, respiratory effort, and blood oxygenation. The type of biosensors used to monitor these parameters for in-laboratory PSG are recommended for use in PMs.

2.2. The sensor to detect apnea is an oronasal thermal sensor and to detect hypopnea is a nasal pressure transducer. Ideally, PMs should use both sensor types.

2.3. Ideally the sensor for identification of respiratory effort is either calibrated or uncalibrated inductance plethysmography.

2.4. The sensor for the detection of blood oxygen is pulse oximetry with the appropriate signal averaging time and accommodation for motion artifact.

3. Methodology for Portable Monitoring

3.1. PM testing should be performed under the auspices of an AASM accredited comprehensive sleep medicine program with policies and procedures for sensor application, scoring, and interpretation of PM. A quality/performance improvement program for PM including inter-scorer reliability must be in place to assure accuracy and reliability.

3.2. An experienced sleep technician, sleep technologist, or appropriately trained healthcare practitioner must perform the application of PM sensors or directly educate the patient in the correct application of sensors.

3.3. PM devices must allow for the display of raw data for manual scoring or editing of automated scoring by a trained and qualified sleep technician/technologist. Evaluation of PM data must include review of the raw data by a board certified sleep specialist or an individual who fulfills the eligibility criteria for the sleep medicine certification examination.

3.4. Scoring criteria should be consistent with the current published AASM standards for scoring of apneas and hypopneas.

3.5. Due to the known rate of false negative PM tests, in laboratory polysomnography should be performed in cases where PM is technically inadequate or fails to establish the diagnosis of OSA in patients with a high pretest probability.

3.6. A follow-up visit with a physician or other appropriately trained and supervised health care provider should be performed on all patients undergoing PM to discuss the results of the test.

3.7. Unattended PM can be used within the parameters specified above in the patient's home.

In 2007, based on their evidence reviews and public comment, CMS approved the use of unsupervised PM sleep studies for the diagnosis of OSA, for the purpose of determining whether treatment with CPAP is warranted. Approval was given for Level 2 and 3 devices. Level 4 devices are acceptable if they record at least three bioparameters.

Identification of Evidence Base

The ideal study for addressing Key Question 4 is a large randomized controlled trial (RCT) that compares crash rates among individuals with OSA who were certified fit-to-drive based on the findings of the current reference standard (PSG) with crash rates among individuals who were certified fit-to-drive based on the findings of an alternative diagnostic. If crash rates are found to be equivalent and the alternative diagnostic was cheaper and more readily available, one would have a compelling argument for utilizing the alternative diagnostic. Unfortunately, no such study exists. Nor, for ethical reasons, is such a study likely to be performed. As a consequence, one must attempt to address Key Question 4 indirectly.

We know from the findings of Key Question 2 that crash risk is directly proportional to the severity of OSA. Consequently, any model, device, or instrument that measures the severity of the disorder (or some a surrogate marker of OSA severity that is known to be associated with crash risk) can potentially be used by a medical examiner to help identify those individuals with OSA who are at an increased risk for a motor vehicle crash.

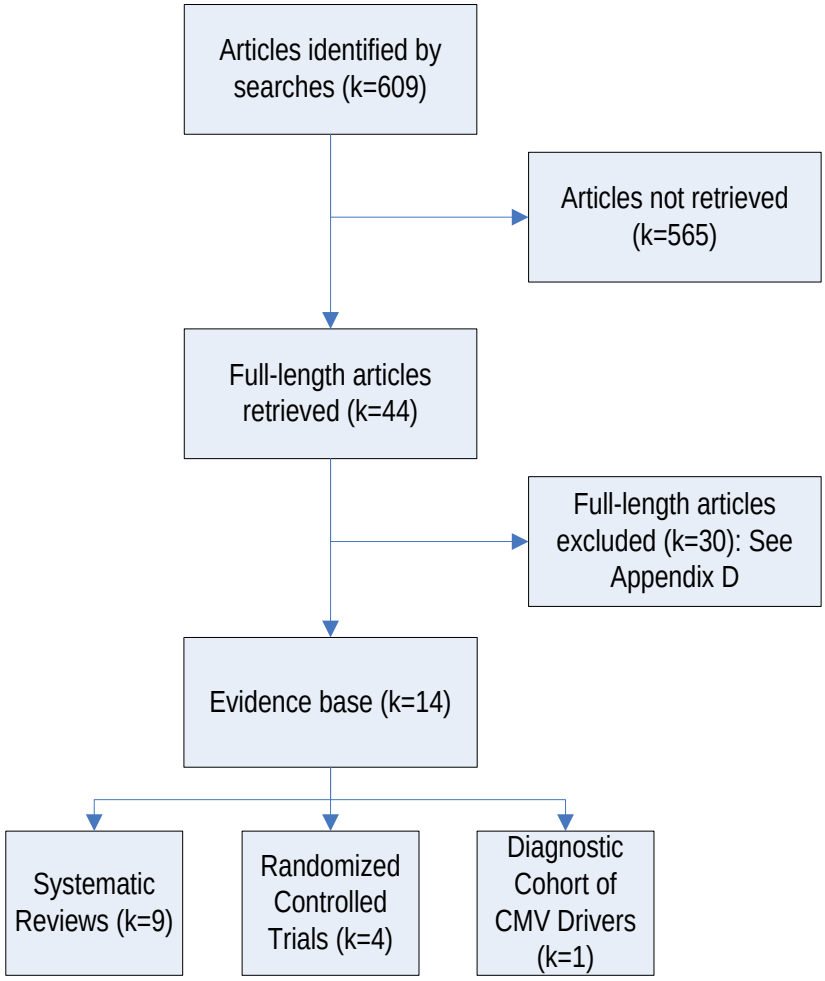
Since publication of the previous evidence report, numerous research studies, including several systematic reviews and RCTs have been conducted comparing PSG and PM devices. This increase in research in this area is partially in response to the CMS decision to approve the use of PM for the diagnosis of OSA.

Given the volume of new research in this area, the evidence base for this updated review is limited to the previously conducted systematic reviews and RCTs. Limiting the update in this manner allowed us to

accurately capture the best possible evidence addressing KQ4 without repeating the work conducted by others.

The identification of the evidence base for Key Question 4 is summarized in Figure 3. Our searches (Appendix A) identified a total of 609 articles that appeared to be relevant to this key question. Following application of the retrieval criteria (Appendix B) for this question, 44 full-length articles were retrieved and read in full. Of these 44 articles, 13 articles met the inclusion criteria (Appendix C) for Key Question 4. One additional study, which did not meet the inclusion criteria, was included because it is relevant to Key Question 4 and was conducted exclusively with CMV drivers. Appendix D lists the 30 articles that were retrieved but then excluded and provides the primary reason for their exclusion.

Figure 3. Development of Evidence Base for Key Question 4



Systematic Reviews Addressing Key Question 4

This subsection provides a brief description of the key attributes of the nine systematic reviews that comprise the evidence base for Key Question 4. Here we discuss applicable information on the

characteristics of the systematic reviews, searches performed, and quality assessments. The characteristics of each included review are presented in Table 12.

Twelve systematic reviews met our inclusion criteria. Several systematic reviews were direct updates of previous reviews. In these cases, we included the most recently conducted review with one exception. Lux and colleagues [113] updated the review conducted by Flemons et al. [3] In their presentation of results, they frequently referred to the results of the previous review (e.g., no new evidence since prior review). Therefore, we included the results of these two reviews together. All systematic reviews compared the effectiveness of portable monitors in diagnosing OSA to PSG. Some reviews also examined the accuracy of automated versus manual scoring of portable monitor results; the functionality of portable monitors (e.g., data loss, malfunction); and cost-benefit of PM devices.

All reviews focused on adults with suspected OSA. The studies included in the reviews were primarily diagnostic cohort studies; two reviews also included RCTs. Four reviews examined Level 2 portable devices, nine examined Level 3 portable devices, and six examined Level 4 portable devices. Three reviews conducted meta-analyses; two additional reviews attempted to conduct meta-analyses but did not due to heterogeneity of the included studies. Severity of OSA was assessed by AHI or RDI and outcomes considered in the reviews were primarily sensitivity, specificity, likelihood ratios, and effect sizes.

Table 13 describes the searches that were performed in each systematic review. Most studies used a combination of database and hand searches (e.g., searches of bibliographies) to identify relevant studies. Three reviews also searched the gray literature. Dates searched ranged from 1966 through January 2008. Inclusion criteria were appropriate and similar across reviews.

Table 12. Characteristics of Included Systematic Reviews

Reference	Year	Population of Interest	Number of Included Studies	Total Number of Participants Included in Studies	Design of Included Studies	Portable Devices Examined	Meta-Analysis Performed?	Assessment of Severity	Outcomes Considered in Review (sens., spec. etc.)
Collop et al. [114]	2007	Adults (≥ 18 years) with suspected OSA	37	4,211	NR	Level 3	No	AHI RDI	Sensitivity Specificity Effect sizes
Flemons et al. [3]	2003	Adults (≥ 18 years) with suspected OSA	51	5,901	Diagnostic Cohort	Level 2 Level 3 Level 4	No, due to heterogeneity	AHI RDI	Sensitivity Specificity Likelihood Ratios
Lux et al. [113]	2004	Adults (≥ 18 years) with suspected OSA	12	2,480	Diagnostic Cohort	Level 3 Level 4	No	AHI RDI	Sensitivity Specificity Likelihood Ratios
Ghegan et al. [115]	2006	Adults (≥ 18 years) with suspected OSA	18	1331	Diagnostic Cohort	Level 2 Level 3 Level 4	Yes	RDI	Effect sizes
Joint Nordic Project [116]	2007	Adult patients with suspected OSA	8	435	Diagnostic Cohort	Level 3	Yes	AHI	Sensitivity Specificity Likelihood Ratios
Thurnheer et al. [117]	2007	Patients with suspected OSA	18	Systematic Review: 784 Original articles: 367	Diagnostic Cohort	Level 3	No	AHI	Sensitivity Specificity Likelihood Ratios False Negative False Positive

Reference	Year	Population of Interest	Number of Included Studies	Total Number of Participants Included in Studies	Design of Included Studies	Portable Devices Examined	Meta-Analysis Performed?	Assessment of Severity	Outcomes Considered in Review (sens., spec. etc.)
Tice [118]	2009	Patients with suspected OSA	5	671	RCT Diagnostic Cohort	Level 3 Level 4	No	AHI RDI	CPAP use Epworth Sleepiness Scale Functional outcomes of sleep questionnaire Sleep apnea quality of life index
Tregear et al. [119]	2007	Adults (≥ 18 years) with suspected OSA	43	4,991	Diagnostic Cohort	Level 2 Level 3 Level 4	Yes	AHI AI RDI PRI AH/TIB CT90 ODI PRRI Peak amplitude HRVI ISI HIS SaO_2	Sensitivity Specificity Positive Predictive Value Negative Predictive Value
Trikalinos et al. [112]	2007	Adults (≥ 18 years) with suspected OSA	95	UC	RCT Diagnostic cohort	Level 2 Level 3 Level 4	No, due to heterogeneity	AHI RDI	Effect size

NR=Not reported

Table 13. Searches performed by Systematic Review Authors

Reference	Year	Databases searched	Hand searches?	Gray literature searched?	Dates of searches	Inclusion criteria	Inclusion criteria appropriate?	Evidence that inclusion criteria not adhered to?
Collop et al. [114]	2007	MEDLINE	No	No	1997 – August 2006 (One additional 2007 paper included)	Subjects ≥ 18 years of age Patient evaluated for OSA Patients had testing with a monitoring device that offered fewer channels (Level 3 devices) than polysomnography Minimum 10 patients	Yes	No
Flemons et al. [3]	2003	MEDLINE	Yes	No	1997 – 2001	Meal/female patients, ages ≥ 18 years, with ANY diagnosis of OSA Study published in English, no race or gender restrictions Portable device used for diagnosis PSG or other acceptable objective test used for the diagnosis of sleep apnea After completion of the study, each analysis group was ≥ 10 subjects No studies in children No review, meta-analyses, case reports, abstracts, letters, and editorials	Yes	No
Lux et al. [113]	2004	MEDLINE The Cochrane Library National Guidelines Clearinghouse International Network of Agencies	Yes	No	From January 2002 – date of last search not reported	Human, both sexes, ages 18 and over, with ANY diagnosis of obstructive sleep apnea (OSA) Portable device used for diagnosis AND polysomnography or other acceptable test used for diagnosis of OSA After completion of study, each analysis	Yes	No

Reference	Year	Databases searched	Hand searches?	Gray literature searched?	Dates of searches	Inclusion criteria	Inclusion criteria appropriate?	Evidence that inclusion criteria not adhered to?
		for Health Technologies Assessment (INAHTA) database				group is greater than or equal to 10 subjects Study published in English		
Ghegan et al. [115]	2006	MEDLINE Cochrane collection	Yes	No	1966 – 2005	Performed both portable and in-laboratory sleep studies in either a simultaneous or sequential fashion in the same group of patients Provided information on at least one of the outcome parameters of interest (FDI, mean low oxygen saturation level, recorded sleep time, quality of sleep study, study cost) for each diagnostic method	Yes	No
Joint Nordic Project [116]	2007	PubMed	Yes	No	Through February 2006	Studies using overnight polysomnography in at least 10 subjects on two separate occasions to detect night-to-night variability of the AHI. Studies comparing portable devices or pulse oximetry or global impression with overnight polysomnography during the same night in at least 10 subjects who had been referred for sleep apnoea recording with the AHI or ODI as outcomes Level 3 or Level 4 devices	Yes	No
Thurnheer et al. [117]	2007	MEDLINE	No	No	Systematic reviews: no	Unattended Level 3 respiratory polygraphy	UC	No

Reference	Year	Databases searched	Hand searches?	Gray literature searched?	Dates of searches	Inclusion criteria	Inclusion criteria appropriate?	Evidence that inclusion criteria not adhered to?
					date limitations Original articles: 2003 – May 2005			
Tice [118]	2009	MEDLINE Cochrane clinical trials Cochrane reviews DARE	Yes	Yes	March 2005 – January 2008	n≥200 for diagnostic studies Had to have home evaluation	UC	No
Tregear et al. [119]	2007	MEDLINE PubMed (pre Medline) EMBASE PsycINFO CINAHL TRIS Cochrane Library	Yes	Yes	Through April 2007	English language ≥ 10 subjects OSA only (e.g., no central apneas) or must present data for OSA separately Sleep studies performed with both facility-based PSG and portable monitors in the same patients, either simultaneously or within 3 months Must report outcomes in terms of sensitivity and specificity of portable monitors related to PSG AI or AHI or allow for calculation	Yes	No
Trikalinos et al. [112]	2007	MEDLINE	Yes	Yes	Through February 2007	English language Medically stable adults with no previous OSA-related surgery	Yes	No

Reference	Year	Databases searched	Hand searches?	Gray literature searched?	Dates of searches	Inclusion criteria	Inclusion criteria appropriate?	Evidence that inclusion criteria not adhered to?
						> 10 subject Sleep studies performed simultaneously or within 3 months Prospective studies		

Quality of Systematic Reviews

The findings of our assessment of the quality of systematic reviews for Key Question 4 are summarized in Table 14. Our assessment found that the quality of reviews ranged from low-to-moderate to high for reporting quality and moderate to high for methodological quality. The primary reasons for lower ratings include no duplicate study selection/data extraction, not providing a list of excluded studies, and inability to assess whether publication status was used as an exclusion criterion.

Table 15 describes the methods used in each systematic review to assess the quality of included studies. Eight out of the nine included reviews assessed the quality of studies included. Six studies used a specific tool to assess quality; the remaining two reviews did not report their methodology for quality assessment. Three reviews also assessed the overall strength of the body of evidence with a particular set of defined criteria.

Table 14. Quality of included systematic reviews*

Reference	Year	Item 1 – Was an 'a priori' design provided?	Item 2 – Were there duplicate study selection and data extraction?	Item 3 – Was a comprehensive literature search performed?	Item 4 – Was the status of publication (i.e., grey literature) used as an exclusion criterion?	Item 5 – Was a list of studies (included and excluded) provided?	Item 6 – Were the characteristics of the included studies provided?	Item 7 – Was the scientific quality of the included studies assessed and reported?	Item 8 – Was the scientific quality of the included studies used appropriately in formulating conclusions?	Item 9 – Were the methods used to combine the findings of studies appropriate?	Item 10 – Was the likelihood of publication bias assessed?	Item 11 –Was conflict of interest identified?	REPORTING QUALITY	METHODOLOGICAL QUALITY
Collop et al. [114]	2007	Y	Y	Y	U	P	Y	P	U	Y	N	N	Low to Mod.	U[a]
Flemons et al. [3]	2003	Y	Y	Y	U	Y	Y	Y	Y	Y	N	N	High	High
Lux et al. [113]	2004	Y	Y	Y	U	N	Y	Y	Y	Y	N	N	Mod. to High	High
Ghegan et al. [115]	2006	Y	NR	Y	U	P	Y	N	NA	Y	Y	N	Mod.	Mod.
Joint Nordic Project [116]	2007	Y	Y	Y	U	Y	Y	Y	Y	Y	N	N	High	High
Thurnheer et al. [117]	2007	Y	NR	Y	U	P	Y	Y	Y	Y	N	N	Mod.	Mod.
Tice [118]	2009	Y	NR	Y	N	P	Y	Y	Y	Y	N	N	Mod.	Mod.
Tregear et al. [119]	2007	Y	N	Y	N	Y	Y	Y	Y	Y	Y	N	High	Mod. to High
Trikalinos et al.	2007	Y	Y	Y	N	P	Y	Y	Y	Y	N	N	Mod. to	High

Reference	Year	Item 1 – Was an 'a priori' design provided?	Item 2 – Were there duplicate study selection and data extraction?	Item 3 – Was a comprehensive literature search performed?	Item 4 – Was the status of publication (i.e., grey literature) used as an exclusion criterion?	Item 5 – Was a list of studies (included and excluded) provided?	Item 6 – Were the characteristics of the included studies provided?	Item 7 – Was the scientific quality of the included studies assessed and reported?	Item 8 – Was the scientific quality of the included studies used appropriately in formulating conclusions?	Item 9 – Were the methods used to combine the findings of studies appropriate?	Item 10 – Was the likelihood of publication bias assessed?	Item 11 –Was conflict of interest identified?	REPORTING QUALITY	METHODOLOGICAL QUALITY
[112]													High	

*Items 1 through 11 are from AMSTAR.

NA=Not Applicable; NR=Not reported; P=partial; U=unclear

[a] Unclear due to low reporting of methodology

Table 15. Methods used to assess quality of studies included in systematic review and strength of body of evidence

Reference	Year	Quality of included studies assessed?	Method used to assess quality of included studies	Strength of body of evidence assessed?	Method used to assess strength of body of evidence
Collop et al. [114]	2007	Yes	NR	No	NA
Flemons et al. [3]	2003	Yes	Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB; *Diagnosis and Screening in Evidence Based Medicine: How to Practice and Teach EBM*. Edinburgh: Churchill Livingstone; 2000;67-93.	Yes	Quantity Quality Consistency
Lux et al. [113]	2004	Yes	Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB; *Diagnosis and Screening in Evidence Based Medicine: How to Practice and Teach EBM*. Edinburgh: Churchill Livingstone; 2000;67-93.	No	NA
Ghegan et al. [115]	2006	No	NA	No	NA
Joint Nordic Project [116]	2007	Yes	Modified QUADAS	Yes	Defined criteria; grading system
Thurnheer et al. [117]	2007	Yes	Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB; *Diagnosis and Screening in Evidence Based Medicine: How to Practice and Teach EBM*. Edinburgh: Churchill Livingstone;	No	NA

Reference	Year	Quality of included studies assessed?	Method used to assess quality of included studies	Strength of body of evidence assessed?	Method used to assess strength of body of evidence
			2000;67-93.		
Tice [118]	2009	Yes	NR	No	NA
Tregear et al. [119]	2007	Yes	ECRI Institute Assessment Tool for Diagnostic Studies	Yes	Tredwell, JT, Tregear, SJ, Reston, JT, Turkelson, CM. (2006). A system for rating the stability and strength of medical evidence. *BMC Medical Research Methodology, 19*, 6-52.
Trikalinos et al. [112]	2007	Yes	Assessment with a set of quality items: Blinding of assessors to results of other test Blinding to clinical information Enrollment of consecutive patients Random order of measurements or simultaneous measurement Proportion of data loss Clear description of the evaluated population	No	NA

NA=Not Applicable; NR=Not reported

Findings

Performance of Portable Monitors in Diagnosing OSA compared to PSG

All nine systematic reviews reported findings on the performance of portable monitors in diagnosing OSA compared to PSG [3, 112-119]. Overall, the reviews found that portable monitors performed reasonably well compared to PSG and may have utility as an alternative to PSG. Most studies examined in these reviews were conducted in populations with moderate to severe sleep apnea with no comorbid conditions so the performance of portable monitors in other populations is largely unknown.

Reviews which examined evidence by level of device found little information related to the performance of Level 2 devices [3, 112-114, 119]. The few studies which were identified related to Level 2 devices indicated that they may be useful as an alternative to PSG (e.g., sensitivities and specificities range from 80%-100% and 90%-100%, respectively); however, more research is needed to confirm this [112, 119].

Most research on the performance of portable monitors has been conducted with Level 3 devices. Most studies show these devices to have good sensitivity and specificity rates compared to PSG, although the rates are not 100% [3, 112, 116, 117, 119]. Specificity tends to be lower than sensitivity indicating that portable monitors may result in higher rates of false positives.

Three reviews examined Level 4 devices separately [3, 112, 119]. These studies indicate that Level 4 devices may also be useful as an alternative to PSG; however more research is needed to confirm this. Level 4 devices tended to perform better if they recorded at least three bioparameters and used lower AHI thresholds.

Table 16. Findings of included systematic reviews related to Performance of Portable Monitors compared to PSG

Reference	Year	Findings
Collop et al. [114]	2007	• Little data on the validity and reliability of Level 2 PM devices exists. Most focuses on Level 3 devices. • PM devices have only been shown to have good specificity and sensitivity in populations evaluated by sleep specialists, considered to be at high risk for OSA based on clinical symptoms and without significant comorbid medical disorders or suspicion of comorbid sleep disorders. It is recommended that PM use be limited to these groups.
Flemons et al. [3]	2003	*Summary* Evidence shows that different monitors show promise for excluding disease, confirming disease or both. Overall, the most consistent, high-quality data were for Level 3 monitors in the attended setting where they had utility to either confirm or exclude sleep apnea in a sleep laboratory population. The number of false results was low in these studies, and the majority of studies were able to find one cutoff RDI that allowed distinction between patients with and without sleep apnea. *PM ability to reduce probability of having abnormal AHI* Level 2 • No evidence supporting PM ability to reduce probability of having abnormal AHI • Included studies had small sample size, non-low LRs, and high false negative rates Level 3 (using flow measured by thermistor) • Some evidence supporting sleep lab-attended PM to reduce probability of having abnormal AHI • Low evidence supporting home-unattended PM to reduce probability of having abnormal AHI • Included studies had low LRs, 4-17% false negative rates Level 4 • Some evidence supporting PM ability to reduce probability of having abnormal AHI • Included studies reported low LRs; eight studies on home-unattended PMs reported 3-37% false-negative rates *PM ability to increase probability of having abnormal AHI*

Reference	Year	Findings
		Level 2 • Some evidence supporting PM ability to increase probability of having abnormal AHI • Included studies had high sensitivities (>90%), high LR (8-40.5) and low false-positive rates (3-11%) Level 3 • Some evidence supporting sleep lab-attended PM ability to increase probability of having abnormal AHI • Included studies on sleep lab-attended PMs had high specificity (>90%), high sensitivity and high LRs • Low evidence supporting home-unattended PM ability to increase probability of having abnormal AHI • Included studies on home-unattended PMs had moderate specificities (58% to 93%), low/moderate LRs (1.8-9) and 2-31% false-positive rates Level 4 • Some evidence supporting PM ability to increase probability of having abnormal AHI • Included studies on sleep lab-attended PM had high LRs, low-to-moderate false positive rates (3-37%) and low data loss (0-11%) • Included studies on home-unattended PM had moderate-to-high LRs; moderate-LR studies had 10-47% false-positive rates *PM ability to both reduce and increase probability of abnormal AHI* Level 2 • Some evidence supporting PM ability to both reduce and increase probability of having abnormal AHI • Included studies had moderate-to-high sensitivities and specificities. One study misclassified 10% of subjects and had high data loss rate (20%) Level 3 • Some evidence supporting PM ability to both reduce and increase probability of having abnormal AHI • Included studies on sleep lab-attended PM had low misclassification rate (5%); home-unattended PM had 5-16% misclassification rate Level 4

Reference	Year	Findings
		• Some evidence supporting PM ability to both reduce and increase probability of having abnormal AHI • Included studies had low misclassification rates (<7%)
Lux et al. [113]	2004	• No new information from previous review • Better quality studies yielded sensitivity and specificity values (or LRs) that provided modest changes in the probability of OSA over the pretest probability
Ghegan et al. [115]	2006	• Home sleep studies provide similar diagnostic information to laboratory PSG but may underestimate sleep apnea severity • RDI values on portable sleep studies were 10% lower on average compared with laboratory studies (OR 0.90; 95% CI 0.87-0.92) • No significant difference in the mean low oxygen saturation on portable versus laboratory studies (OR 1.0; 95% CI 0.94-1.10)
Joint Nordic Project [116]	2007	o Manually scored portable devices including airflow, respiratory movements and pulse oximetry during one night of sleep have high sensitivity and specificity to identify a pathologic AHI compared with PSG (Evidence Grade 1) o Automatic scoring of the results of portable devices has high sensitivity and identifies most patients with a pathologic AHI, but specificity is low (Evidence Grade 1) o Pulse oximetry with results from the oxygen desaturation index is insufficient to identify a pathologic AHI and there is a high risk that patients with sleep apnoea syndrome will be incorrectly classified as normal (Evidence Grade 1)
Thurnheer et al. [117]	2007	• With a mean pre-test probability of 64% for OSA, the post-test probability after a negative result ranged from 8% (negative likelihood ratio of 0.05) to 23% (negative likelihood ratio of 0.20). • The post-test probability after a positive result was within a range of 90% (positive likelihood ratio of 23.8) to 98% (positive likelihood ration of 5.7).
Tice [118]	2009	• Included studies demonstrated equivalence of an ambulatory strategy incorporating portable home monitoring to PSG for the following outcomes: AHI on CPAP, sleepiness, and quality of life or functional status o Remmers Sleep Recorder and WatchPat 100 are just as useful as PSG for identifying and instituting treatment in patients at very high risk for OSA when used in conjunction with ESS
Tregear et al. [119]	2007	*Summary* • While no portable sleep monitoring system was as accurate as the reference standard (none had a sensitivity and specificity of 100%), analyses found that the diagnostic performance characteristics of most portable systems were reasonable. That is, the vast majority of available systems could differentiate individuals with OSA from those without and they could differentiate individuals with severe OSA

Reference	Year	Findings
		from those with mild-to-moderate disease better that would be expected by chance alone o A number of portable sleep monitoring systems, though not as accurate as the current reference standard (laboratory PSG) do offer an alternative method by which the severity of OSA may be assessed in a large number of individuals at a relatively low cost o Whether these systems are accurate enough to be considered as acceptable alternative to the current reference standard for stratifying individuals by OSA severity for the purposes of making decisions about the fitness of an individual to drive a CMV is not clear. *Level 2 Monitors* • Included studies had a range of sensitivities of 80-100% and a range of specificities of 90-100%. • Meta-analytic pooling of the results of included studies yielded summary sensitivities of 80-100%, and summary specificities of 90-100%. Summary estimates generally increased at higher AHI thresholds. *Level 3 Monitors* • Included studies had a range of sensitivities of 33.3-100% (higher sensitivities were generally observed at lower AHI thresholds), and a range of specificities of 25-100% • Meta-analytic pooling of the results of included studies yielded summary sensitivities of 44-98.8%, and summary specificities of 81-95.4%. Summary sensitivities generally decreased at higher AHI thresholds. *Level 4 Monitors* • Included studies had a range of sensitivities of 28-100% (higher sensitivities were generally observed at lower AHI thresholds), and a range of specificities of 19-100% • Meta-analytic pooling of the results of included studies yielded summary sensitivities of 64.6-92.1%, and summary specificities of 83.7-95.4%. Summary sensitivities generally decreased at higher AHI thresholds.
Trikalinos et al. [112]	2007	• Based on limited data, Level 2 monitors may identify people with AHI suggestive of OSA with high positive likelihood ratios and low negative likelihood ratios • Level 3 monitors may have the ability to predict AHI suggestive of OSA with high positive likelihood ratios and low negative likelihood ratios for various AHI cutoffs in laboratory-based PSG, especially with manual scoring is employed • Studies of Level 4 monitors that record at least three bioparameters showed 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

Reference	Year	Findings
		setting compared to studies in the specialized sleep laboratory

Automated versus Manual Scoring of Events

Five systematic reviews [3, 112-114, 116] compared the results of manual and automated scoring of PM testing with the results from PSG. All reviews found that manual scoring was more comparable to PSG than automated. This finding was consistent for all levels of portable monitoring. The current AASM guidelines incorporate these findings, specifically stating:

"PM devices must allow for the display of raw data for manual scoring or editing of automated scoring by a trained and qualified sleep technician/technologist. Evaluation of PM data must include review of the raw data by a board certified sleep specialist or an individual who fulfills the eligibility criteria for the sleep medicine certification examination." [114]

Table 17. Findings of Included Systematic Reviews related to Automated vs. Manual Scoring of Portable Monitor Results

Reference	Year	Findings
Collop et al. [114]	2007	• Nine studies provided evidence related to the comparability of automated versus manual scoring; all showed manual scoring to be superior.
Flemons et al. [3]	2003	• Five studies compared automated vs. manual scoring of portable monitor results; all five studies found manual scoring to be superior.
Lux et al. [113]	2004	• Results for automated methods of scoring respiratory events appear to provide less agreement with PSG results than do manual methods. That is, for portable monitors with data recordings that could be scored either manually or with automated algorithms (or both), manual scoring produced results with better concordance with PSG results.
Joint Nordic Project [116]	2007	• Manual scoring of portable devices (n = 6) compared with polysomnography during the same night in hospital had high pooled sensitivity of 0.93 (95% ci 0.89–0.97) and high specificity of 0.92 (95% ci 0.87–0.96) (Evidence Grade 1) • Automatic scoring of portable devices (n = 3) compared with polysomnography had a pooled sensitivity of 0.92 (95% ci 0.83–0.97) without heterogeneity and a pooled specificity of 0.85 (95% ci 0.73–0.93) with heterogeneity (p = 0.010) (Evidence Grade 1) • Automated scoring programs cannot score sleep time and it is unclear whether these programs can differentiate obstructive from central apneas.
Trikalinos et al. [112]	2007	• 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.

Functionality of Portable Monitors

Five systematic reviews [3, 112-115] examined the functionality of portable monitors, specifically data loss during testing. All five reviews found that portable monitors resulted in more data loss compared to PSG. Two reviews also included studies which compared home and laboratory based portable monitors. Trikalinos et al. [112] found that data corruption was higher for portable monitors in the home setting; however, the results presented by Flemons et al. [3] are less clear. Recent reviews [112, 114] indicated that newer devices may be better equipped to prevent these errors through the use of built in alert systems.

AASM have incorporated these findings into their guidelines, specifically stating:

"An experienced sleep technician, sleep technologist, or appropriately trained healthcare practitioner must perform the application of PM sensors or directly educate the patient in the correct application of sensors." [114]

Table 18. Findings of Included Systematic Reviews related to Functionality of Portable Monitors

Reference	Year	Findings
Collop et al. [114]	2007	The evidence review of portable monitors reported data loss of 3%-18% for Level 3 monitors and 7%-10% for oxygen saturation measurements in Level 4 monitors. The subsequent AHRQ review noted inadequate or missing data precluding adequate interpretation reported in 13%-20% of studies for Level 3 monitors. Golpe and coworkers reported data loss that prevented interpretation in 7% of studies in which a technologist applied the sensors as compared to 33% in which the patient applied the sensors independently at home. A more recent study found that 5.6% of patients had more than 20% of time in bed with absent or inadequate airflow. For Level 4 devices used in the home, data loss was reported to be between 11% and 16%. This study also reported on the use of the PM in the laboratory with technologist application of the sensors resulting in only 3% data loss. A new Level 4 device provides an audible alarm if the device comes off or need adjustment. This approach resulted in only 2% of studies with insufficient data.
Flemons et al. [3]	2003	***Failure of PM*** Level 2 • Three of four included studies reported data loss. Level 3 • Three studies of home-unattended PM reported data loss ranging from 3% to 18% • Two studies of sleep lab-attended PM reported data loss of 3.3% and 9% respectively Level 4 • Four studies of home-unattended PM reported data loss ranging from 7% to 10% • Eleven studies of sleep lab-attended PM reported data loss ranging from 2% to 14%
Lux et al. [113]	2004	Reported data loss in the home studies considered for this update ranged from 3 percent to 33 percent (in a subgroup). Moreover, at the upper end of this data loss range, many experts doing systematic reviews of clinical literature would probably regard the studies as being of only poor quality and perhaps not give them further consideration. Only one home study *directly* compared the data loss rate between hook-up for the portable equipment by technicians and that by patients; 11 the investigators reported a 7-percent loss for technician hook-up and 33-percent loss for patient hook-up. In the study using PAT technology, 14 three of 28 (11 percent) of initial home studies set up by the patients were "rejected" whereas only three of 102 (3 percent) of studies done in the laboratory with equipment hooked up by a technician were "rejected." Thus, although only a limited amount of evidence in the reports reviewed addresses this issue, data loss appears to be greater when the patient performs the hook-up of the equipment.
Ghegan et al. [115]	2006	• Recorded sleep time was significantly higher by 13% for laboratory compared with portable studies (OR 0.87; 95% CI 0.86-0.89) • Portable studies were significantly more likely to give a poor recording when compared with laboratory examinations (p=.0001)

Reference	Year	Findings
Trikalinos et al. [112]	2007	• 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 • Signal loss was more often observed in home studies; indirect evidence suggests that these errors can possibly be prevented through the use of built in alert systems • 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

Cost-Benefit of Portable Monitors

Five systematic reviews [3, 113-115, 117] reported on the potential cost-benefit of portable monitors compared to PSG. Most reviews demonstrated cost saving of portable monitors even when accounting for higher rates of data loss.

Table 19. Findings of included systematic reviews related to Cost-Benefit of Portable Monitors

Reference	Year	Findings
Collop et al. [114]	2007	Although the actual purchase cost of PM devices has decreased substantially over the past decade, the total health care costs of evaluating and treating suspected sleep apnea using PM have not been adequately compared to the costs using PSG. In the prior practice parameter, of 12 studies utilizing 3 or more recording channels (Level 3), only 2 actually reported costs of PM in the home. Twenty-two percent to 33% cost savings were described when compared to in-lab PSG. In the 35 studies previously reviewed, 5 attended, 2 unattended and 1 mixed population in-home protocols were described. Cost savings were universally reported, but there was significant variability in study design, pretest probability for OSA and threshold level (i.e., RDI) for the diagnosis of sleep apnea. Among the more recent studies, Dingli and coworkers reported a 42% savings including technician time, supplies, and equipment depreciation if patients went straight from PM to CPAP. However, they had 24% initial and 12% later PM failure rates which must be included in such analysis. Bachour and coworkers evaluated an esophageal monitor with flow and oximetry compared to in-lab polysomnography, each followed by CPAP titration. Some savings were noted, but the analysis did not include costs incurred from reevaluation of the 40% of false negative studies. Regardless of the number of channels recorded, none of the studies previously or currently reviewed address costs relative to the popular use of split study protocols (an initial baseline evaluation in the laboratory followed by nasal CPAP titration in appropriately selected patients). Furthermore, costs of treatment are often not compared, such as those incurred when auto-adjusting positive airway pressure (APAP) or empiric CPAP home treatment protocols versus the standard 2 night baseline and CPAP in-laboratory titration studies are used.
Flemons et al. [3]	2003	*Cost-benefit analyses of PM* Level 2 • Of the 2 included studies, one reported PM at half the cost of PSG Level 3 • Of 12 included studies, 6 didn't mention cost, 3 assumed PM cost less than PSG and 1 reported time-saved in using PM over PSG (115 mins vs. 225 mins, respectively) • Two studies of home-unattended PM reported 66-78% savings when using PM over PSG Level 4 • Of 35 included studies, 8 reported some form of cost-benefit analyses • Five included studies reported on how use of PM could completely exclude the need for PSG • Two included studies reported cost-savings of 44% and 77% when using PM over PSG • One included study reported a savings of £62.00 per patient when using PM over PSG
Lux et al. [113]	2004	Three included studies examined the cost-benefit of portable monitors. These studies found: • Portable monitors reduced diagnostic costs by 42% if those in the diagnostic categories went straight to CPAP titration or no further

Reference	Year	Findings
		investigation with only those in the possible OSAHS and failed home study groups proceeding to PSG. • Home studies with a technician setup of the equipment were less expensive (because of the high percentage of faulty studies with patient own setup of sleep recording devices). • As a screening tool for the diagnosis of OSA, pulse oximetry is cost effective and shows substantial accuracy.
Ghegan et al. [115]	2006	• Cost of home studies ranged from 35% to 88% lower than laboratory studies across a number of countries
Thurnheer et al. [117]	2007	• Assuming 5000 PMs and PSGs per year in Switzerland, the use of PM over PSG in patients with high clinical suspicion of OSA led to 71.6% savings (2,225,385 SFr vs. 7,830,630 SFr). This result takes into account 3.5% of inconclusive PMs that required confirmatory PSG.

Randomized Controlled Trials Related to Key Question 4

Four RCTs were identified that met the inclusion criteria for Key Question 4. Three studies [120-122] examined differences in clinical outcomes after CPAP therapy in subjects diagnosed using portable monitors compared to subjects diagnosed using PSG. The fourth study [123] used a clinical algorithm (i.e., Epworth Sleepiness Scale, Sleep Apnea Clinical Score, and portable home monitoring) to identify patients with a high probability of OSA; patients were then randomized to receive a confirmatory PSG and CPAP titration PSG or autotitrating CPAP. Two studies addressed Level 3 portable sleep monitors and two addressed Level 4 sleep monitors (Table 20). The primary characteristics of the four included studies are presented in Table 21.

Table 20. Evidence Base for Key Question 4

Reference	Year	Study Location	Country
LEVEL 3 Sleep Monitors			
Berry et al. [120]	2008	Malcom Randall Veterans Administration Medical Center (VAMC)	USA
Skomro et al. [121]	2010	University of Saskatchewan, Saskatoon	Canada
LEVEL 4 Sleep Monitors			
Mulgrew et al. [123]	2007	University of British Columbia, Vancouver, British Columbia	Canada
Whitelaw et al. [122]	2005	University of Calgary, Calgary, Alberta	Canada

Table 21. Key Study Design Characteristics of Studies that Address Key Question 4

Reference	Year	Portable System	Study Design	N (% male)	Setting	Assessment Of Severity	Reference Standard (PSG)	Participants	Consecutive patients?	Parallel Enrollment?
LEVEL 3 SLEEP MONITORS										
Berry et al. [120]	2008	WatchPAT100	RCT	PM: 53 (87% male) PSG: 53 (88.7% male)	1) Lab 2) Home	AHI	Full night/7 parameters	Referrals to Malcom Randall VAMC for diagnosis of suspected OSA	Y	Y
Skomro et al. [121]	2010	Embletta	RCT	PM: 44 (67% male) PSG: 45 (68% male)	1) Lab 2) Home	AHI RDI	Split-night and Full night/ 5 parameters	Adult outpatients with suspected OSA referred to participating sleep medicine physicians at a tertiary outpatient sleep disorders clinic	NR	Y
LEVEL 4 SLEEP MONITORS										
Mulgrew et al. [123]	2007	AutoSet Spirit Ohmeda Biox 3400	RCT	PM: 33 (79% male) PSG: 35 (75% male)	1) Lab 2) Home	AHI	Full night/2parameters	Referrals from catchment area of Sleep Disorders Program at University of British Columbia Hospital, Vancouver, British Columbia, for assessment of suspected obstructive sleep apnea	Y	Y
Whitelaw et al. [122]	2005	Snoresat	RCT	PM: 156 PSG: 132 Gender NR	1) Lab 2) Home	AHI	Full night/ 2 parameters	Patients were a randomly selected subset of consecutive eligible patients referred by family doctors to a sleep center. The inclusion criterion was a history suggesting OSA in association with somnolence or fatigue.	Y	Y

AI=Apnea index; AHI =Apnea-hypopnea index; RDI=Respiratory disturbance index; ODI =Oxygen desaturation Index; PRI =Portable respiratory index; PPRI =Pulse rate rise index; AH/TIB =apnea + hypopnea per hour of time in bed; CT90 =Cumulative time spend below a saturation of 90%; SaO_2 ≥Oxygen saturation

Quality of Included Studies

The findings of our assessment of the quality of the included studies are presented in Table 22. Studies were assessed using the ECRI Institute Quality Scale I: Controlled Trials. Our assessment found the quality of the included studies to be in the moderate and moderate-to-high range. The main potential for bias found in these studies included a lack of blinding among participants and practitioners.

Table 22. Quality of the Studies that Address Key Question 4

Reference	Year	Items																								Quality Category
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
LEVEL 3 SLEEP MONITORS																											
Berry et al. [120]	2008	Y	NR	N	Y	Y	Y	Y	N	Y	Y	Y	Y	N	NA	N	NR	NR	Y	Y	Y	Y	Y	Y	N	Y	Moderate
Skomro et al. [121]	2010	Y	Y	N	Y	N	Y	Y	N	Y	Y	N	Y	N	NA	N	NR	NR	Y	Y	Y	Y	Y	Y	Y	Y	Moderate
LEVEL 4 SLEEP MONITORS																											
Mulgrew et al. [123]	2007	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	NA	N	NR	N	Y	Y	Y	Y	Y	Y	N	Y	Moderate – to – High
Whitelaw et al. [122]	2005	Y	Y	N	Y	N	Y	Y	N	N	Y	N	Y	N	NA	N	NR	NR	Y	Y	Y	Y	Y	Y	Y	Y	Moderate

NR=not reported; NA =not applicable

Generalizability of Evidence to Target Population

Important characteristics of the individuals included in the studies that address Key Question 4 are summarized in Table 23. The generalizability of the individuals enrolled in the four included studies to CMV drivers is unclear. No studies provided information about the occupation or driving experience of the participants making it difficult to generalize on the basis of employment or driving exposure. CMV drivers in the United States tend to be older (over 40 years) and male which is consistent with the samples included in these studies.

Table 23. Individuals in Studies that Address Key Question 4

Reference	Year	N=	Participants	Mean Age (years)	PSG?/ number of patients with OSA	Severity (mean AHI)	% male	% CMV drivers	Generalizability to CMV population
LEVEL 3 Sleep Monitors									
Berry et al. [120]	2008	PM: 53 PSG: 53	Referrals	PM: 51.9 ± 1.7 PSG: 55.1 ± 1.5	PM: 50 PSG: 49	PM: 29.2 ± 2.3 (TST) PSG: 36.8 ± 4.8 (TST) 27.8 ± 3.3 (TRT)	PM: 87% PSG: 88.7%	NR	Unknown
Skomro et al. [121]	2010	PM: 44 PSG: 45	Referrals	PM: 49.6 ± 10.3 PSG: 47.8 ± 11.3	89	PM: 22.3 PM/PSG: 28.8 PSG: 31.7 PSG/PM: 25.1	PM: 67% PSG: 68%	NR	Unknown
LEVEL 4 Sleep Monitors									
Mulgrew et al. [123]	2007	PM: 33 PSG: 35	Referrals	PM: 55±10 PSG: 52 ± 11	PM: 31 PSG: 30	Median AHI: PM: 2.5 episodes/hr PSG: 3.2 episodes/hr	PM: 79% PSG: 75%	NR	Unknown
Whitelaw et al. [122]	2005	PM: 156 PSG: 132	Referrals	PM: 46.9 ± 10.2 PSG: 46.9 ± 9.7	NR	Median AHI: PM: 16.6 PSG: 26	NR	NR	Unknown

NR =Not reported

Findings

Our searches identified four studies that compared the use of PM and PSG for diagnosis and treatment of OSA. Studies examined a variety of clinical outcomes including difference in ESS scores, Sleep Apnea Quality of Life Index (SAQLI) scores, and residual AHI. Table 24 and Table 25 summarize the individual study findings followed by a brief description of the major findings of these studies.

Table 24. Results of Included RCTs

Reference	Year	N=	Portable system assessed	Setting	Assessment of severity	Threshold	Overall Results
Level 3 Sleep Monitors							
Berry et al. [120]	2008	PM: 53 PSG: 53	WatchPAT100	Lab Home	AHI	AHI ≥5	The AHI in the PM-APAP group was 29.2 ± 2.3/h and in the PSG group was 36.8 ± 4.8/h (P = NS). Patients with an AHI ≥ 5 were offered CPAP treatment. Those accepting treatment (PM-APAP 45, PSG 43) were begun on CPAP using identical devices at similar mean pressures (11.2 ± 0.4 versus 10.9 ± 0.5 cm H_2O). At a clinic visit 6 weeks after starting CPAP, 40 patients in the PM-APAP group (78.4% of those with OSA and 88.8% started on CPAP) and 39 in the PSG arm (81.2% of those with OSA and 90.6% of those started on CPAP) were using CPAP treatment (P = NS). The mean nightly adherence (PM-APAP: 5.20 ± 0.28 versus PSG: 5.25 ± 0.38 h/night), decrease in Epworth Sleepiness Scale score (–6.50 ± 0.71 versus –6.97 ± 0.73), improvement in the global Functional Outcome of Sleep Questionnaire score (3.10 ± 0.05 versus 3.31 ± 0.52), and CPAP satisfaction did not differ between the groups.
Skomro et al. [121]	2010	PM: 44 PSG: 45	Embletta	Lab Home	AHI RDI	AHI≥5 RDI>5	After 4 weeks of CPAP therapy, there were no significant differences in: ESS (PSG 6.4 ± 3.8 vs home monitoring [HM] 6.5 ± 3.8, *P* = .71) PSQI (PSG 5.4 ± 3.1 vs HM 6.2 ± 3.4, *P* = .30) SAQLI (PSG 4.5 ± 1.1 vs HM 4.6 ± 1.1, *P* = .85) SF-36 vitality (PSG 62.2 ± 23.3 vs HM 64.1 ± 18.4, *P* = .79) SF-36 HM (PSG 84.0 ± 10.4 vs HM 81.3 ± 14.9, *P*= .39) BP (PSG 129/84 ± 11/0 vs HM 125/81 ± 13/9, *P* = .121) There was no difference in CPAP adherence (PSG 5.6 ± 1.7 h/night vs HM 5.4 ± 1.0 h/night, *P* = .49).
Level 4 Sleep Monitors							
Mulgrew et al. [123]	2007	PM: 33 PSG: 35	Autoset Spirit Ohmeda Biox 3400	Lab Home	AHI	AHI ≥5	The PSG and ambulatory groups had similar median BMI (38 kg/m2), age (55 years), ESS score (14 points), and respiratory disturbance index (31 episodes of respiratory disturbance/h). After 3 months, there were no differences in the primary outcome, AHI on CPAP (median, 3.2 vs. 2.5; difference, 0.8/h [95% CI, -0.9 to 2.3]) (*P* =0.31), between the PSG and

Reference	Year	N=	Portable system assessed	Setting	Assessment of severity	Threshold	Overall Results
							ambulatory groups, or in the secondary outcomes, ESS score, Sleep Apnea Quality of Life Index, and CPAP. Adherence to CPAP therapy was better in the ambulatory group than in the PSG group (median, 5.4 vs. 6.0; difference, -1.12 h/night [CI, -2.0 to 0.2]) (*P*= 0.021).
Whitelaw et al. [122]	2005	PM: 156 PSG: 132	Snoresat	Lab Home	AHI	NR	Measured outcomes of treatment in the two groups of patients were not statistically significantly different. Mean increases in SAQLI were 0.92 in the polysomnography group and 0.82 in the home monitoring group (p=0.50; 95% confidence interval for the difference in SAQLI:-0.40 to +0.20). Forty-four percent of the polysomnography and 40% of home monitor patients improved (p = 0.47; 95% confidence interval for the difference: -4 to +17%). Mean decreases in Epworth scale were 4.0 for polysomnography and 3.4 for home monitoring (p=0.27). The mean RDI on treatment was 5.7 for polysomnography and 4.2 for home monitoring (p=0.06). Mean scores and improvements in scores in domains of SF-36 were not significantly different.

AHI =Apnea-hypopnea index; RDI=Respiratory disturbance index; ODI ≥Oxygen desaturation Index

Table 25. Results of Specific Outcomes Reported Across RCTs

Reference	Year	Followup Time	OSA (%)	CPAP Attempted (%)	CPAP continued (%)	CPAP Hrs/night	Nights > 4 hrs use, %	Residual AHI	Change ESS	Change SAQLI
LEVEL 3 SLEEP MONITORS										
Berry et al.* [120]	2008	6 weeks	PM: 96% PSG: 91% p=NS	PM: 85% PSG: 81% p=NS	PM: 75% PSG: 74% p=NS	PM: 5.2 ± 0.28 PSG: 5.3 ± 0.38 p=NS	PM: 72 ± 29.09 PSG: 67 ± 39.97 p=NS	PM: 3.5 ± 1.90 PSG: 5.3 ± 4.37 p=NS	PM: -6.5 ± 4.49 PSG: -7.0 ± 4.56 p=NS	PM: NR PSG: NR
Skomro et al. [121]	2010	4 weeks	87%	PM: 73% PSG: 82% p=NR	PM: 64.7% PSG: 73% p=NR	PM: 5.4 ± 1.0 PSG: 5.6 ± 1.7 p=NS	PM: 88 PSG: 89 p=NS	PM: NR PSG: NR	PM: 13.0 ± 3.6, 6.5 ± 3.8 PSG: 12.5 ± 5.0, 6.4 ± 3.8 p=NS	PM: 4.6 ± 1.1 PSG: 4.5 ± 1.1 p=NS
LEVEL 4 SLEEP MONITORS										
Mulgrew et al.[a] [123]	2007	3 months	89%	PM: 100% PSG: 94% p=NR	PM: 94% PSG: 86% p=NR	PM: 6.0 (5.1-7.1) PSG: 5.4 (3.7-6.4) p=0.021	PM: NR PSG: NR	PM: 2.5 (0.9 – 10.1) PSG: 3.2 (1.7 – 8.4) p=NS	PM: -8.0 (-4.0 - -12.0) PSG: -10.0 (-6.0.- -12.0) p=NS	PM: 1.9 (1.1 – 3.0) PSG: 2.2 (1.2 – 3.4) p=NS
Whitelaw et al. [122]	2005	4 weeks	NA	PM: 100% PSG: 100% p=NS	PM: 82% PSG: 82% p=NS	PM: 3.3 PSG: 3.8 p=NS	PM: 56 PSG: 62 p=NS	PM: 4.2 PSG: 5.7 p= 0.06	PM: -3.4 PSG: -4.0 p=NS	PM: 0.82 PSG: 0.92 p=NS

*Standard Deviations calculated by MANILA

[a] Outcome data presented as medians and interquartile ranges due to non-normality of data; effect sizes calculated by authors using Hodges-Lehmann estimate for difference in medians.

NS = not significant; NR = not reported

Berry 2008

Berry et al. [120] conducted a randomized trial that examined the use of PM and unattended autotitrating positive airway pressure (APAP) for selecting an effective continuous positive airway pressure (CPAP) as a means to diagnose OSA. Watch PAT100 was the PM device used in the study. The authors compared the PM pathway with another pathway; PSG. Patients referred to the Malcom Randall Veterans Administration Medical Center (VAMC) for suspected OSA were considered for inclusion in the study. Patients were included if they had significant daytime sleepiness (ESS≥12) and at least two of the following: habitual loud snoring, witnessed apnea/gasping, or treatment for hypertension. Patients were excluded if they lived more than 200 miles from VAMC, had congestive heart failure, used nocturnal oxygen, had chronic obstructive pulmonary disease, awake hypercapnia, neuromuscular disease, cataplexy, restless legs syndrome, had prior diagnosis or treatment of a sleep disorder, used potent narcotics, or suspected other causes for sleepiness (shift work).

The subjects in the PSG group underwent CPAP titration if their AHI was ≥ 10/hour during the first two hours of monitoring. Those patients with an AHI ≥ 5/hour underwent CPAP titration on another night. Patients in the PM arm applied the Watch PAT100 device themselves at home before going to sleep. In the morning, they brought the device back to the medical center where the AHI was calculated. Patients with an AHI ≥ 5/hour were diagnosed with OSA, fitted with a mask and sent home with an APAP device. The unattended APAP and attended CPAP groups were then offered treatment with CPAP. The same device (REMstar Pro with C-Flex and heated humidity) was used for those who accepted CPAP treatment. After their initial study, subject with an AHI < 5/hour were re-evaluated using the alternate diagnostic test. After six weeks, the mean nightly use in hours of CPAP among patients still using CPAP was assessed. The Epworth Sleepiness Scale (ESS) and Functional Outcomes of Sleep Questionnaire (FOSQ) were measured at baseline and at the clinic evaluation after six weeks on CPAP. Satisfaction with CPAP and the residual AHI were also analyzed.

A total of 106 patients were randomized; 53 to the PM group and 53 to the PSG group. The subjects' mean age was about 53 years old. Eighty-eight percent (88%) were male and the average BMI was about 34 kg/m2. In both groups, the mean ESS was slightly over 16. At baseline and at the six week CPAP evaluation, there were no significant differences between groups. The mean nightly adherence (PM: 5.20 ± 0.28 vs. PSG: 5.25 ± 0.38 hour/night), decrease in ESS score (–6.50 ± 0.71 versus –6.97 ± 0.73), improvement in the global FOSQ score (3.10 ± 0.05 versus 3.31 ± 0.52). While not significant, patients in the home diagnosis group were slightly more satisfied, (6.5 vs. 5.6). In the PM group, 51 out of 53 (96%) were diagnosed with OSA, 45 out of 53 (85%) accepted setup of CPAP, and 40 out of 53 (75%) completed follow-up through six-weeks. In the PSG group, 48 out of 53 (91%) were diagnosed with OSA, 43 out of 53 (81%) accepted setup of CPAP, and 39 out of 53 (74%) completed follow-up through six-weeks. Overall, 93% of patients had OSA.

Skomro 2010

Skomro et al. [121] conducted a randomized trial that compared the subjective sleepiness, sleep quality, quality of life, blood pressure, and continuous positive airway pressure (CPAP) adherence after four

weeks of CPAP treatment in patients who were diagnosed and treated at home for OSA with those who underwent in-laboratory PSG. Those who were treated at home used the Embletta device. Participants were adult outpatients with suspected OSA referred to the participating sleep medicine physicians at a tertiary outpatient sleep disorders clinic. They were considered eligible if they were older than 18 years, had at least two symptoms of OSA, and lived within an hour drive from the center. Participants were excluded if they had heart or respiratory failure, clinical symptoms of another sleep disorder, a safety-sensitive occupation, used hypnotics, had upper airway surgery, CPAP or oxygen therapy, were pregnant or were unable to provided informed consent.

Patients randomized to the PM arm underwent one (1) night of level three testing with Embletta. The patients then underwent an in-laboratory overnight PSG after completing the home testing and prior to the auto-CPAP. If AHI was ≥15, CPAP was applied during the PSG. Subjects with an AHI was >5 but <15 repeated the in-laboratory PSG with CPAP titration. Those with a respiratory disturbance index (RDI) >5 were diagnosed with OSA. All of these subjects (RDI>5) were offered auto-CPAP therapy for one (1) week followed by fixed pressure CPAP based on 95% pressure derived from the auto-CPAP device. Patients with an RDI <5 were withdrawn from the study. Patients in the PSG arm completed an in-laboratory overnight PSG followed by one night of PM. CPAP titration was performed either during the split-night PSG (if AHI was >15) or during a second in-laboratory PSG (if the AHI was >5, but <15). Those with an AHI of >5 were diagnosed with OSA and offered CPAP therapy at the pressure obtained in the sleep laboratory.

The outcomes were measured using the Epworth Sleepiness Scale (ESS), the Pittsburgh Sleep Quality Index (PSQI), the Calgary Sleep Apnea Quality of Life Index (SAQLI), and the 36-Item Short-Form Health Survey (SF-36). Once CPAP therapy began, all patients were evaluated at week 1 with ESS and CPAP compliance, then at week 4 with ESS, SAQLI, SF-36, PSQI, arterial BP, and CPAP compliance. Daytime somnolence (ESS) at 4 weeks was the primary outcome. A total of 102 patients were randomized. The subjects' mean age was 47 years and 62% percent were male. Average body mass index (BMI) was 32 kg/m² and mean ESS score was 12.5. After the HM and PSG sleep testing, 89 subjects were considered to have OSA and subsequently prescribed CPAP therapy. Ten subjects rejected the CPAP therapy. For those adhering to the CPAP therapy, after 4 weeks of there were no significant differences in ESS (PSG 6.4 vs. HM 6.5), PSQI (PSG 5.4 vs. HM 6.2), SAQLI (PSG 4.5 vs. HM 4.6), SF-36 vitality (PSG 62.2 vs. HM 64.1), and BP (PSG 129/84 vs. HM 125/81). There was no difference in CPAP adherence (PSG 5.6 h/night vs. HM 5.4 h/night).

Skomro's study results indicate that when compared to PM diagnosis and therapy, PSG diagnosis and treatment for OSA does not lead to superior 4-week outcomes.

Mulgrew 2007

Mulgrew et al. [123] conducted a randomized trial that compared the results of a diagnostic strategy based on PM to laboratory PSG in high-risk obstructive sleep apnea (OSA) patients. Eligible participants were those consecutively referred to a tertiary sleep center who had suspected OSA. Patients were included if they had a high pretest probability of OSA (AHI>15 episodes/hour), were medically stable and

not taking any sedative medications. Exclusion criteria included prior treatment for OSA, unwillingness to use CPAP, pregnancy, unstable angina, abnormal spirometry, unwillingness to sign the informed consent and significant psychiatric illness.

Subjects were randomized to PSG or the ambulatory diagnostic arm at home (auto-CPAP, home oximetry pressure set, home oximetry adjust and fix CPAP, treat fixed CPAP at monthly clinic visits) based on their ESS and respiratory disturbance index (RDI) scores. The ResMed AutoSet Spirit autotitrating CPAP machine was used by all patients. Patients randomized to PSG took part in an overnight in-laboratory assessment. CPAP was determined based on standard protocol and no subsequent adjustment to the fixed CPAP occurred. For the ambulatory group, after 1 week the device software was examined to gather data on CPAP, mask leak, residual respiratory events and use of the device. On days 6 and 13, overnight oximetry was performed using the Ohmeda Biox3400. The following day, patients went back to the sleep center and if respiratory events or oxygen de-saturation had been observed, the 95th% pressure or the fixed CPAP was increased. On the 14th day, the final pressure was set and did not change for the remainder of the study.

The primary outcome was AHI on CPAP after three months of treatment. At baseline and after three months of therapy, patients completed the Epworth Sleepiness Scale (ESS) (decreased score is less sleepy) and the Sleep Apnea Quality of Life Index (SAQLI) (increased score reflects improved quality of life). A total of 68 patients had an RDI>15 per hour and were randomized to PSG or ambulatory groups. Thirty-five were randomized to PSG and 33 to the ambulatory arm. Both groups had similar median BMI (38 kg/m^2), age (55 years), ESS score (14 points), and RDI (31 episodes of respiratory disturbance/ hours). After three months there was no difference in AHI on CPAP (PSG 3.2/hour; ambulatory 2.5/hour, p=0.31). Adherence to CPAP therapy was better in the ambulatory group than in the PSG group (median, 5.4 vs. 6.0; difference, -1.12 h/night [CI, -2.0 to 0.2]) (p =0.021). There were also no significant differences in ESS (PSG -10 points versus ambulatory -8 points, p=0.26) or SAQLI (2.2 versus 1.9, p=0.69). CPAP compliance was slightly worse in the PSG group (5.4 vs. 6.0 hours per night, p=0.021).

The authors concluded that the PSG-based diagnostic approach offers no advantage over the ambulatory approach in the initial management of patients with a high probability of sleep apnea and ambulatory CPAP compliance may be superior. Patients in most need of treatment should use the ambulatory method when access to PSG is inadequate.

Whitelaw 2005

The primary objective of PM devices and PSG is to identify which patients have OSA symptoms that will benefit from treatment. To compare the accuracy of PM with PSG, Whitelaw et al. [122] randomized patients referred to a sleep center to undergo PSG or home monitoring. Eligible participants were consecutively referred, having a history of OSA symptoms along with somnolence or fatigue. Individuals were excluded if they had a non-respiratory sleep disorder as the main reason for their referral, lack of important daytime symptoms, significant comorbidities, and considerable physiologic consequences of OSA. PSG was a standard full-night diagnostic study. The PM device used was Snoresat. The primary

outcome measure was successful treatment defined as an increase greater than 1.0 in Sleep Apnea Quality of Life Index (SAQLI).

Based on clinical data and test results, sleep specialists estimated the likelihood of success of treatment as greater than 50% or less than 50%. All patients then had a 4 week treatment with auto-adjusting continuous positive airway pressure (CPAP). Overall, 307 patients were randomized with 288 patients completing the four weeks of CPAP and final SAQLI questionnaire. The PSG group consisted of 132 patients and the PM group, 156. There were no significant differences between the two groups. Mean age was 47 years, body mass index (BMI) was 32 kg/m2, their neck circumference was 41 cm and their score on the standard Epworth Sleepiness Scale (ESS) was 11.6. Overall, 42% of patients met criteria for successful treatment (≥1 point increase in SAQLI). The correct prediction rate was 61% for patients who had PSG and 64% for patients who had home monitoring (p=0.72). The ability to predict successful response to CPAP using PSG was no more than 7% better than the rate for using PM.

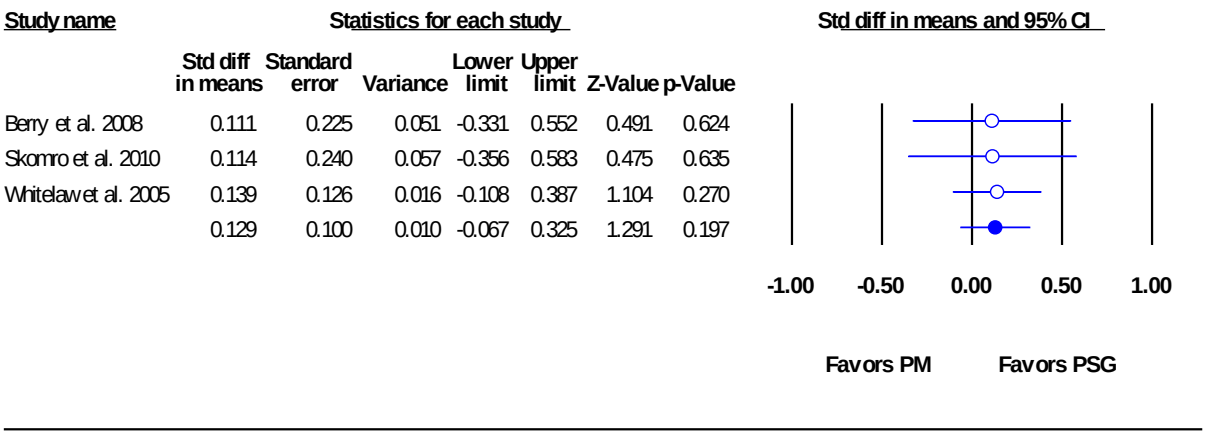
Overall, the findings showed no significant advantage to using PSG over PM for the purpose of identifying patients benefit from OSA treatment. The authors provide four reasons for poor accuracy in predicting successful treatment with PSG or home monitoring: (1) there are various unknowns (low quality of life due to symptoms, symptoms due to OSA, patient tolerates CPAP, and benefits of treatment outweigh side effects) that are considered when predicting successful treatment and with so many uncertainties accuracy is, not surprisingly, poor; (2) placebo effect or regression to the mean effect correct evaluation; (3) it is very difficult to make successful predictions for those patients judged to be close to the 50% success definition - they fall above or below the 50% by chance; (4) a one point improvement in the SAQLI may not be a good metric for successful treatment.

Meta-Analysis of Epworth Sleepiness Scale (ESS) Score

All four RCTs [120-123] presented data on ESS scores in a manner that allowed for calculation of effect sizes (i.e., standardized mean differences) and pooling of data. Because the design of the Mulgrew et al. [123] study differed from the other three studies it was not included in this meta-analysis.

The findings of these analyses are presented below (Figure 4). A test for heterogeneity found the studies to be homogenous (I^2=0.000). We therefore conducted a fixed effects meta-analysis to determine whether ESS scores after CPAP treatment differ between portable monitoring and PSG groups. The summary standardized difference in means was 0.129 (95% CI: -0.067, 0.335; p=0.325), suggesting a trend toward slightly better scores among the PSG group, although this difference between groups was not statistically significant.

Figure 4. Meta-analysis of ESS Score – Portable Monitoring vs. PSG



Portable Monitoring Devices versus PSG in a Commercial Driving Population

An additional study was included in this report even though it did not meet inclusion criteria because of its focus on PM and PSG exclusively among CMV drivers. Below we provide a brief summary of this study.

Study Summary

Watkins et al. [124] utilized a prospective case series study design to evaluate the accuracy of PM (the RUSleeping RTS device) compared with PSG, in a subset of CMV)drivers screening positive for OSA using Consensus Conference Criteria for the screening of OSA among CMV drivers. [35, 97]

The study enrolled 346 newly-hired CMV drivers who screened positive for OSA by the Consensus Conference Criteria; a positive screen was defined as meeting one or more of the following criteria:

1. Sleep history (snoring, excessive daytime sleepiness, witnessed apneas)
2. History of MVA likely related to sleep disturbance such as run off road, at-fault, rear-end collision
3. Previous OSA diagnosis, prior polysomnography (PSG) with AHI >5, or reported continuous positive airway pressure (CPAP) use
4. Epworth sleepiness scale (ESS) with a score >10
5. Sleeping in the examination room
6. Two or more of the following:
    a. Body mass index ≥35 kg/m2
    b. Neck circumference (NC) > 17 inches in men, 16 inches in women
    c. Hypertension (new, uncontrolled, or requiring more than or equal to two medications for control)

The PM used in the study was the RUSleeping RTS device: a single-channel device designed to measure airflow (nasal pressure) by a nasal cannula and thus to detect the rate of apnea events (AHI) that occurred during the time the device was turned on. The drivers were instructed by the study clinician to turn the device on when they went to bed and to turn it off as soon as they awoke the next morning. Drivers used the PM for one night only. Drivers underwent PSG testing at a certified community sleep center under the supervision of sleep specialists approximately 6-8 weeks after PM.

One-hundred-and-nine (32%) subjects screened positive for OSA according to the Consensus Conference Criteria; 68 (62%) of these subjects completed PM and a further 34 (50%) of these subjects also underwent PSG. The mean PSG AHI was 18.5 (range = 0.8-117), while the mean RUSleeping AHI was 14.7 (range = 9-68.7).

Forty-six of the 57 (80%) drivers who underwent PM appeared to have some degree of OSA (AHI > 5); 26 of the 34 drivers who underwent PSG (74%) were confirmed positive for some degree of OSA (AHI > 5) and 8 drivers had severe OSA (AHI > 30). When comparing PM to PSG at a definition of OSA as an AHI >15, the PPV is 0.64 with a 0.87 NPV (see Table 26). The sensitivity is 0.70 and specificity is 0.83. The

positive likelihood ratio for comparing RUSleeping AHI of >15 with a PSG AHI >15 is 4.20, suggesting that those with a high pretest probability by Consensus Conference Criteria at an AHI >15 would likely have PSG-confirmed OSA. The calculated area under the receiver-operator curve was 0.885.

Investigators found a statistically significant relationship between continuous RUSleeping AHI and continuous PSG AHI ($p = 0.0004$). Pearson correlation coefficients of the RUSleeping device with AHI demonstrated that at AHI >15, the RUSleeping device is statistically significantly correlated with PSG, $R = 0.57$, $p < 0.001$.

Table 26. Diagnostic Performance of RUSleeping portable monitoring device

Comparison of RUSleeping™ AHI to PSG AHI >15

Category	Sensitivity	Specificity	PPV	NPV	LR (+)	LR (–)
AHI >5	1.00	0.42	0.42	1.00	1.17	0.00
AHI >10	1.00	0.71	0.71	1.00	3.43	0.00
AHI >15	0.70	0.83	0.83	0.87	4.20	0.36
AHI >20	0.70	0.83	0.83	0.87	4.20	0.36
AHI >30	0.43	0.96	0.96	0.87	11.57	0.59

Key: AHI = Apnea/Hypopnea Index; LR (+) = Likelihood Ratio positive; LR (-) = Likelihood Ratio Negative; NPV = Negative Predictive Value; PSG = Polysomnography; PPV = Positive Predictive Value

There were several limitations to this study. The sample size was relatively small. The portable monitoring device was worn only for a single night and did not take into account night-tonight variability associated with persons with mild to moderated OSA. Subjects may have been susceptible to a "first night effect," which is the effect of the environment and sleep recording. To most accurately validate the portable device, the equipment would have ideally been worn simultaneously with PSG testing. Driver compliance with the portable device was a major issue: the 33 drivers who did not complete both PM and PSG testing could not be evaluated.

Comparison of Study Findings with those of Key Question 4 in Original Sleep Apnea Review

The RUSleeping RTS device is classified as a Level 4 Sleep Monitor, measuring only airflow (nasal pressure). One may use the RUSleeping RTS device at home (versus at a sleep lab). The original OSA review [119] examined the diagnostic performances of five Level 4, home-based portable monitors (the OxiFlow (OS), the Aposcreen I, the Watch_Pat 100 & two oximeters) [125-129]. The diagnostic performances of these five portable monitors, as well as that of the RUSleeping RTS device, are presented in Table 27.

Table 27. Diagnostic Performances of Level 4, home-based portable monitors

Reference		Baltzan et al. (2000) [127]	Golpe et al. (2002) [126]	Pittman et al. (2004) [125]	Ryan et al. (1995) [128]	Series et al. (1993) [129]	Watkins et al. (2009) [124]
At AHI/RDI ≥ 5	Sensitivity	90%	NR	NR	NR	NR	100%
	Specificity	32%	NR	NR	NR	NR	42%
	PPV	NR	NR	NR	NR	NR	42%
	NPV	NR	NR	NR	NR	NR	100%
At AHI/RDI ≥ 10	Sensitivity	55%	91%	82%	NR	98.2%	100%
	Specificity	88%	81%	100%	NR	47.4%	71%
	PPV	NR	NR	NR	NR	61.4%	71%
	NPV	NR	NR	NR	NR	96.9%	100%
At AHI/RDI ≥ 15	Sensitivity	34%	NR	96%	32%	NR	70%
	Specificity	94%	NR	100%	100%	NR	83%
	PPV	NR	NR	NR	NR	NR	83%
	NPV	NR	NR	NR	NR	NR	87%
At AHI/RDI ≥ 20	Sensitivity	31%	NR	80%	NR	NR	70%
	Specificity	97%	NR	89%	NR	NR	83%
	PPV	NR	NR	NR	NR	NR	83%
	NPV	NR	NR	NR	NR	NR	87%
At AHI/RDI ≥ 30	Sensitivity	7%	NR	92%	NR	NR	43%
	Specificity	100%	NR	82%	NR	NR	96%
	PPV	NR	NR	NR	NR	NR	96%
	NPV	NR	NR	NR	NR	NR	87%

Key: AHI = Apnea/Hypopnea Index; NR = Not Reported; RDI = Respiratory Disturbance Index

The monitors investigated by Golpe et al. [126] and Series et al. [129] were studied at an AHI/RDI of ≥ 10 only, and predicted OSA with a sensitivity of 91% and 98.2%, and a specificity of 81% and 47.4%, respectively – values lower than the 100% sensitivity and 71% specificity of the Watkins et al. [124] monitor at an AHI of ≥ 10. The monitor investigated by Ryan et al. [128] was studied only at an AHI of ≥ 15; its sensitivity and specificity was 32% and 100% respectively, compared to the 70% sensitivity and 83% specificity of the Watkins et al. [124] monitor at AHI ≥ 15. The sensitivity of the Pittman et al. [125] monitor zigzagged between 80% and 92% from an RDI ≥ 10 to ≥ 30; high values when compared to the sensitivities of the Watkins et al. [124] monitor (100% at AHI ≥ 5 to 43% at AHI ≥ 30). Specificities of the Pittman et al. [125] monitor ranged from 100% at RDI ≥ 10 to 82% at RDI ≥ 30, comparable to the range of specificities of the Watkins et al. [124] monitor (71% at AHI ≥ 10 to 96% at AHI ≥ 30). The sensitivity of the Baltzan et al. [127] monitor ranged from 90% at RDI ≥ 5 to 7% at AHI ≥ 30, while its specificity ranged from 32% to 100% for the same RDI thresholds. The sensitivity range of the Baltzan et al. [127] monitor is wider than that of the Watkins et al. [124] monitor (100% at AHI ≥ 100% to 43% at AHI ≥ 30); its specificity range is more comparable (42% at AHI ≥ 5 to 96% at AHI ≥ 30).

Of the five studies that investigated home-based, Level 4 portable monitors in the original OSA review, only the study by Series et al. [129] reported positive and negative predictive values (PPV & NPV), and only at an AHI ≥ 10. The Series et al. [129] monitor had a PPV of 61.4% (vs. a PPV of 71% reported by

Watkins et al. at AHI ≥ 10) and an NPV of 96.9% (vs. a NPV of 100% reported by Watkins et al. [124] at AHI ≥ 10).

Summary of Findings

Fourteen articles provided evidence to inform the conclusions drawn from this updated systematic review examining the performance of PM devices in the diagnosis of OSA compared to the current gold standard, PSG. The findings of our analyses of these 14 studies are summarized below:

The findings of this updated systematic review support our previous findings that a number of portable sleep monitoring systems, though not as accurate as the current reference standard (PSG) do offer an alternative method by which the severity of PSA may be assessed in a large number of individuals at a relatively low cost.

Nine systematic reviews examining the performance of portable monitors in diagnosing OSA compared to PSG found that portable monitors performed reasonably well compared to PSG though none were as accurate (i.e., no PM device has a sensitivity and specificity of 100%). These reviews found that the majority of PM devices could differentiate individuals with OSA from those without and could differentiate individuals with severe OSA from those with mild-to-moderate OSA. Evidence was strongest for Level 3 PM devices for which more research has been conducted. Evidence does indicate that Level 2 and 4 devices show some utility, more research is needed to confirm these findings. Other findings from the systematic reviews indicate that manual scoring of PM devices provide results more consistent with PSG than automated scoring of PM devices; PM devices tend to result in more data loss than PSG although newer devices with built-in alert systems may help reduce these errors, and; PM tends to be associated with higher cost savings over PSG even when accounting for higher rates of data loss.

RCTs examining differences in clinical outcomes after CPAP treatment based OSA diagnosis with PM versus PSG, also support the utility of PM devices in the diagnosis of OSA. A variety of clinical outcomes were assessed across the four studies including AHI, sleepiness, quality of life, and functional and physical health. Very few differences were found between PM and PSG groups on any of these outcomes.

Three RCTs provided information in a manner that allowed us to conduct a meta-analysis. Specifically, we conducted a fixed effects meta-analysis to determine whether ESS scores after CPAP treatment differed between PM and PSG groups. The summary standardized difference in means was 0.129 (95% CI: -0.067, 0.335; p=0.325), suggesting a trend toward slightly better scores among the PSG group, although this difference was not statistically significant.

Bibliography

1. Young, T., P.E. Peppard, and D.J. Gottlieb, *Epidemiology of obstructive sleep apnea: a population health perspective.* Am J Respir Crit Care Med, 2002. **165**(9): p. 1217-39.
2. Caples, S.M., A.S. Gami, and V.K. Somers, *Obstructive sleep apnea.* Ann Intern Med, 2005. **142**(3): p. 187-97.
3. Flemons, W.W., 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): p. 1543-79.
4. Worsnop, C., R. Pierce, and R.D. McEvoy, *Obstructive sleep apnoea.* Aust N Z J Med, 1998. **28**(4): p. 421-7.
5. McNicholas, W.T., *Sleep apnoea and driving risk. European Respiratory Society Task Force on "Public Health and Medicolegal Implications of Sleep Apnoea".* Eur Respir J, 1999. **13**(6): p. 1225-7.
6. Treadwell, J.R., et al., *A system for rating the stability and strength of medical evidence.* BMC Med Res Methodol, 2006. **6**: p. 52.
7. Shadish, W.R. and C.K. Haddock, *Combining estimates of effect size*, in *The handbook of research synthsis*, H. Cooper and L.V. Hedges, Editors. 1994, Russell Sage Foundation: New Your. p. 261-277.
8. Parmar, M.K., V. Torri, and L. Stewart, *Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints.* Stat Med, 1998. **17**(24): p. 2815-34.
9. Hedges, L.V., *Fixed effect models*, in *The handbook of research synthesis*, H. Cooper and L.V. Hedges, Editors. 1994, Russell Sage Foundation: New York. p. 285-299.
10. Hedges, L.V. and J.L. Vevea, *Fixed- and random-effects models in meta-analysis.* Psychol Methods, 1998. **3**(4): p. 486-504.
11. Raudenbush, S.W., *Random effects models*, in *The handbook of research synthesis*, H. Cooper and L.V. Hedges, Editors. 1994, Russell Sage Foundation: New York. p. 301-321.
12. Sutton, A.J., et al., eds. *Methods for meta-analysis in medical researchq*. 2001, John Wiley & Sons. 274.
13. Fleiss, J.L., *Measures of effect size for categorical data*, in *The handbook of research synthesis*, H. Cooper and L.V. Hedges, Editors. 1994, Russell Sage Foundation: New York. p. 245-260.
14. Moses, L.E., D. Shapiro, and B. Littenberg, *Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations.* Stat Med, 1993. **12**(14): p. 1293-316.
15. Littenberg, B. and L.E. Moses, *Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method.* Med Decis Making, 1993. **13**(4): p. 313-21.
16. Gavaghan, D.J., R.A. Moore, and H.J. McQuay, *An evaluation of homogeneity tests in meta-analyses in pain using simulations of individual patient data.* Pain, 2000. **85**(3): p. 415-24.
17. Takkouche, B., C. Cadarso-Suarez, and D. Spiegelman, *Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis.* Am J Epidemiol, 1999. **150**(2): p. 206-15.
18. Higgins, J.P. and S.G. Thompson, *Quantifying heterogeneity in a meta-analysis.* Stat Med, 2002. **21**(11): p. 1539-58.
19. Greenhouse, J.B. and S. Iyengar, *Sensitivity analysis and diagnostics*, in *The handbook of research synthesis*, H. Cooper and L.V. Hedges, Editors. 1995, Russell Sage FOundation: New York. p. 383-409.
20. Petitti, D.B., *Approaches to heterogeneity in meta-analysis.* Stat Med, 2001. **20**(23): p. 3625-33.
21. Higgins, J.P., et al., *Measuring inconsistency in meta-analyses.* BMJ, 2003. **327**(7414): p. 557-60.

22. van Houwelingen, H.C., L.R. Arends, and T. Stijnen, *Advanced methods in meta-analysis: multivariate approach and meta-regression.* Stat Med, 2002. **21**(4): p. 589-624.
23. Thompson, S.G. and J.P. Higgins, *How should meta-regression analyses be undertaken and interpreted?* Stat Med, 2002. **21**(11): p. 1559-73.
24. Higgins, J.P. and S.G. Thompson, *Controlling the risk of spurious findings from meta-regression.* Stat Med, 2004. **23**(11): p. 1663-82.
25. Duval, S. and R. Tweedie, *Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis.* Biometrics, 2000. **56**(2): p. 455-63.
26. Sutton, A.J., et al., *Empirical assessment of effect of publication bias on meta-analyses.* BMJ, 2000. **320**(7249): p. 1574-7.
27. Cooper, H. and L.V. Hedges, eds. *The handbook of research synthesis*. 1994, Russell Sage FOundation: New York. 573.
28. Ancoli-Israel, S., et al., *Expert Panel Recommendations: Obstructive Sleep Apnea and Commercial Motor Vehicle Driver Safety*. 2008.
29. *Prevalence of Sleep Apnea Among Commercial Truck Drivers*. [cited 2011 October 20]; Available from: http://www.atri-online.org/research/results/sleep_apnea_white_paper.pdf.
30. Cistulli, P.A. and R.R. Grunstein, *Medical devices for the diagnosis and treatment of obstructive sleep apnea.* Expert Rev Med Devices, 2005. **2**(6): p. 749-63.
31. *Evidence Report: Obstructive Sleep Apnea and Commercial Motor Vehicle Driver Safety (Comprehensive Review)*. 2007, The ECRI Institute.
32. Hartenbaum, N., et al., *Sleep apnea and commercial motor vehicle operators: Statement from the joint task force of the American College of Chest Physicians, the American College of Occupational and Environmental Medicine, and the National Sleep Foundation.* Chest, 2006. **130**(3): p. 902-5.
33. Chervin, R.D., et al., *Cost-utility of three approaches to the diagnosis of sleep apnea: polysomnography, home testing, and empirical therapy.* Ann Intern Med, 1999. **130**(6): p. 496-505.
34. Gurubhagavatula, I., et al., *Occupational screening for obstructive sleep apnea in commercial drivers.* Am J Respir Crit Care Med, 2004. **170**(4): p. 371-6.
35. Parks, P., et al., *Screening for obstructive sleep apnea during commercial driver medical examinations.* J Occup Environ Med, 2009. **51**(3): p. 275-82.
36. Young, T., et al., *Predictors of sleep-disordered breathing in community-dwelling adults: the Sleep Heart Health Study.* Arch Intern Med, 2002. **162**(8): p. 893-900.
37. Ancoli-Israel, S., *Epidemiology of sleep disorders.* Clin Geriatr Med, 1989. **5**(2): p. 347-62.
38. Young, T., J. Skatrud, and P.E. Peppard, *Risk factors for obstructive sleep apnea in adults.* JAMA, 2004. **291**(16): p. 2013-6.
39. Dixon, J.B., et al., *Daytime sleepiness in the obese: not as simple as obstructive sleep apnea.* Obesity (Silver Spring), 2007. **15**(10): p. 2504-11.
40. Friedman, M., et al., *Clinical predictors of obstructive sleep apnea.* Laryngoscope, 1999. **109**(12): p. 1901-7.
41. Hoffstein, V. and S. Mateika, *Differences in abdominal and neck circumferences in patients with and without obstructive sleep apnoea.* Eur Respir J, 1992. **5**(4): p. 377-81.
42. Cherniack, N.S., *Sleep apnea and its causes.* J Clin Invest, 1984. **73**(6): p. 1501-6.
43. Harman, E.M., J.W. Wynne, and A.J. Block, *The effect of weight loss on sleep-disordered breathing and oxygen desaturation in morbidly obese men.* Chest, 1982. **82**(3): p. 291-4.

44. Phillips, B., et al., *Sleep apnea: prevalence of risk factors in a general population.* South Med J, 1989. **82**(9): p. 1090-2.
45. Stradling, J.R. and J.H. Crosby, *Predictors and prevalence of obstructive sleep apnoea and snoring in 1001 middle aged men.* Thorax, 1991. **46**(2): p. 85-90.
46. Chervin, R.D. and C. Guilleminault, *Obstructive sleep apnea and related disorders.* Neurol Clin, 1996. **14**(3): p. 583-609.
47. Namyslowski, G., et al., *Sleep study in patients with overweight and obesity.* J Physiol Pharmacol, 2005. **56 Suppl 6**: p. 59-65.
48. Kyzer, S. and I. Charuzi, *Obstructive sleep apnea in the obese.* World J Surg, 1998. **22**(9): p. 998-1001.
49. Schafer, H., et al., *Body fat distribution, serum leptin, and cardiovascular risk factors in men with obstructive sleep apnea.* Chest, 2002. **122**(3): p. 829-39.
50. Katz, I., et al., *Do patients with obstructive sleep apnea have thick necks?* Am Rev Respir Dis, 1990. **141**(5 Pt 1): p. 1228-31.
51. Chaouat, A., et al., *Association of chronic obstructive pulmonary disease and sleep apnea syndrome.* Am J Respir Crit Care Med, 1995. **151**(1): p. 82-6.
52. Ferguson, K.A., et al., *The relationship between obesity and craniofacial structure in obstructive sleep apnea.* Chest, 1995. **108**(2): p. 375-81.
53. Nishimura, Y., et al., *Obesity and obstructive sleep apnea syndrome.* Acta Otolaryngol Suppl, 2003(550): p. 22-4.
54. Pillar, G., et al., *Predictive value of specific risk factors, symptoms and signs, in diagnosing obstructive sleep apnoea and its severity.* J Sleep Res, 1994. **3**(4): p. 241-244.
55. Ward Flemons, W. and W.T. McNicholas, *Clinical prediction of the sleep apnea syndrome.* Sleep Med Rev, 1997. **1**(1): p. 19-32.
56. Davies, R.J., N.J. Ali, and J.R. Stradling, *Neck circumference and other clinical features in the diagnosis of the obstructive sleep apnoea syndrome.* Thorax, 1992. **47**(2): p. 101-5.
57. Ogretmenoglu, O., et al., *Body fat composition: a predictive factor for obstructive sleep apnea.* Laryngoscope, 2005. **115**(8): p. 1493-8.
58. Santaolalla Montoya, F., et al., *The predictive value of clinical and epidemiological parameters in the identification of patients with obstructive sleep apnoea (OSA): a clinical prediction algorithm in the evaluation of OSA.* Eur Arch Otorhinolaryngol, 2007. **264**(6): p. 637-43.
59. Cao, J., et al., *[The primary diagnostic significance of the epworth sleepiness scale in patients with obstructive sleep apnea syndrome].* Zhonghua Jie He He Hu Xi Za Zhi, 2002. **25**(3): p. 154-5.
60. Johns, M.W., *A new method for measuring daytime sleepiness: the Epworth sleepiness scale.* Sleep, 1991. **14**(6): p. 540-5.
61. Johns, M.W., *Reliability and factor analysis of the Epworth Sleepiness Scale.* Sleep, 1992. **15**(4): p. 376-81.
62. Uribe Echevarria, E.M., et al., *[Epworth drowsiness scale value in obstructive sleep apnea syndrome].* Medicina (B Aires), 2000. **60**(6): p. 902-6.
63. Sher, A.E., et al., *Predictive value of Muller maneuver in selection of patients for uvulopalatopharyngoplasty.* Laryngoscope, 1985. **95**(12): p. 1483-7.
64. Schwab, R.J. and A.N. Goldberg, *Upper airway assessment: radiographic and other imaging techniques.* Otolaryngol Clin North Am, 1998. **31**(6): p. 931-68.
65. Rivlin, J., et al., *Upper airway morphology in patients with idiopathic obstructive sleep apnea.* Am Rev Respir Dis, 1984. **129**(3): p. 355-60.

66. Roux, F., C. D'Ambrosio, and V. Mohsenin, *Sleep-related breathing disorders and cardiovascular disease.* Am J Med, 2000. **108**(5): p. 396-402.
67. Kapuniai, L.E., et al., *Identifying sleep apnea from self-reports.* Sleep, 1988. **11**(5): p. 430-6.
68. Bliwise, D.L., J.C. Nekich, and W.C. Dement, *Relative validity of self-reported snoring as a symptom of sleep apnea in a sleep clinic population.* Chest, 1991. **99**(3): p. 600-8.
69. Grunstein, R., et al., *Snoring and sleep apnoea in men: association with central obesity and hypertension.* Int J Obes Relat Metab Disord, 1993. **17**(9): p. 533-40.
70. Shinohara, E., et al., *Visceral fat accumulation as an important risk factor for obstructive sleep apnoea syndrome in obese subjects.* J Intern Med, 1997. **241**(1): p. 11-8.
71. Chen, N.H., et al., *A two-tier screening model using quality-of-life measures and pulse oximetry to screen adults with sleep-disordered breathing.* Sleep Breath, 2010.
72. Crocker, B.D., et al., *Estimation of the probability of disturbed breathing during sleep before a sleep study.* Am Rev Respir Dis, 1990. **142**(1): p. 14-8.
73. Dixon, J.B., L.M. Schachter, and P.E. O'Brien, *Predicting sleep apnea and excessive day sleepiness in the severely obese: indicators for polysomnography.* Chest, 2003. **123**(4): p. 1134-41.
74. Khoo, S.M., et al., *Diagnostic characteristics of clinical prediction models for obstructive sleep apnea in different clinic populations.* Sleep Breath, 2010.
75. Morris, L.G., et al., *Rapid risk stratification for obstructive sleep apnea, based on snoring severity and body mass index.* Otolaryngol Head Neck Surg, 2008. **139**(5): p. 615-8.
76. Pradhan, P.S., R.E. Gliklich, and J. Winkelman, *Screening for obstructive sleep apnea in patients presenting for snoring surgery.* Laryngoscope, 1996. **106**(11): p. 1393-7.
77. Rauscher, H., W. Popp, and H. Zwick, *Model for investigating snorers with suspected sleep apnoea.* Thorax, 1993. **48**(3): p. 275-9.
78. Rowley, J.A., L.S. Aboussouan, and M.S. Badr, *The use of clinical prediction formulas in the evaluation of obstructive sleep apnea.* Sleep, 2000. **23**(7): p. 929-38.
79. Sharma, S.K., et al., *A stepped approach for prediction of obstructive sleep apnea in overtly asymptomatic obese subjects: a hospital based study.* Sleep Med, 2004. **5**(4): p. 351-7.
80. Sharma, S.K., et al., *Prediction of obstructive sleep apnea in patients presenting to a tertiary care center.* Sleep Breath, 2006. **10**(3): p. 147-54.
81. Viner, S., J.P. Szalai, and V. Hoffstein, *Are history and physical examination a good screening test for sleep apnea?* Ann Intern Med, 1991. **115**(5): p. 356-9.
82. Whiting, P., et al., *The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews.* BMC Med Res Methodol, 2003. **3**: p. 25.
83. Greenland, S., *Quality scores are useless and potentially misleading: Reply to "Re: A critical look at some popular analytic methods".* American Journal of Epidemiology, 1994. **140**(3): p. 300-302.
84. Kuna, S.T., *Portable-monitor testing: an alternative strategy for managing patients with obstructive sleep apnea.* Respir Care. **55**(9): p. 1196-215.
85. Portier, F., et al., *Evaluation of home versus laboratory polysomnography in the diagnosis of sleep apnea syndrome.* Am J Respir Crit Care Med, 2000. **162**(3 Pt 1): p. 814-8.
86. Le Bon O, H.G., Tecco J, et al., *Mild to moderate sleep respiratory events: one negative night may not be enough.* Chest, 2000. **118**(2): p. 353-359.
87. Adams, N., et al., *Relation of measures of sleep-disordered breathing to neuropsychological functioning.* Am J Respir Crit Care Med, 2001. **163**(7): p. 1626-31.
88. Bennett, L.S., et al., *Health status in obstructive sleep apnea: relationship with sleep fragmentation and daytine sleepiness, and effects of continuous positive airway pressure treatment.* Am J Respir Crit Care Med, 1999. **159**(6): p. 1884-90.

89. Bennett, L.S., et al., *Sleep fragmentation indices as predictors of daytime sleepiness and nCPAP response in obstructive sleep apnea.* Am J Respir Crit Care Med, 1998. **158**(3): p. 778-86.
90. Gottlieb, D.J., et al., *Relation of sleepiness to respiratory disturbance index: the Sleep Heart Health Study.* Am J Respir Crit Care Med, 1999. **159**(2): p. 502-7.
91. Kingshott, R.N., et al., *Predictors of improvements in daytime function outcomes with CPAP therapy.* Am J Respir Crit Care Med, 2000. **161**(3 Pt 1): p. 866-71.
92. Olson, L.G., et al., *A community study of snoring and sleep-disordered breathing. Prevalence.* Am J Respir Crit Care Med, 1995. **152**(2): p. 711-6.
93. Barbe, F., et al., *Treatment with continuous positive airway pressure is not effective in patients with sleep apnea but no daytime sleepiness. a randomized, controlled trial.* Ann Intern Med, 2001. **134**(11): p. 1015-23.
94. Roure, N., et al., *Daytime sleepiness and polysomnography in obstructive sleep apnea patients.* Sleep Med, 2008. **9**(7): p. 727-31.
95. Wiegand, D.M., R.J. Hanowski, and S.E. McDonald, *Commercial drivers' health: a naturalistic study of body mass index, fatigue, and involvement in safety-critical events.* Traffic Inj Prev, 2009. **10**(6): p. 573-9.
96. Smith, B. and B.A. Phillips, *Truckers drive their own assessment for obstructive sleep apnea: a collaborative approach to online self-assessment for obstructive sleep apnea.* J Clin Sleep Med. **7**(3): p. 241-5.
97. Talmage, J.B., et al., *Consensus criteria for screening commercial drivers for obstructive sleep apnea: evidence of efficacy.* J Occup Environ Med, 2008. **50**(3): p. 324-9.
98. Polit, D.F. and C.T. Beck, *Nursing Research: Principles and Methods*. 7th ed. 2004, Philadelphia, PA: Lippincott Williams & Wilkins.
99. Hartenbaum, N., et al., *Sleep apnea and commercial motor vehicle operators: statement from the joint Task Force of the American College of Chest Physicians, American College of Occupational and Environmental Medicine, and the National Sleep Foundation.* J Occup Environ Med, 2006. **48**(9 Suppl): p. S4-37.
100. Kiely, J.L., et al., *Comparison of a limited computerized diagnostic system (ResCare Autoset) with polysomnography in the diagnosis of obstructive sleep apnoea syndrome.* Eur Respir J, 1996. **9**(11): p. 2360-4.
101. Reichert, J.A., et al., *Comparison of the NovaSom QSG, a new sleep apnea home-diagnostic system, and polysomnography.* Sleep Med, 2003. **4**(3): p. 213-8.
102. Parra, O., et al., *Should patients with sleep apnoea/hypopnoea syndrome be diagnosed and managed on the basis of home sleep studies?* Eur Respir J, 1997. **10**(8): p. 1720-4.
103. Verse, T., et al., *Validation of the POLY-MESAM seven-channel ambulatory recording unit.* Chest, 2000. **117**(6): p. 1613-8.
104. Fietze, I., et al., *Automated analysis of data is inferior to visual analysis of ambulatory sleep apnea monitoring.* Respiration, 2002. **69**(3): p. 235-41.
105. White, D.P., et al., *Assessment of accuracy and analysis time of a novel device to monitor sleep and breathing in the home.* Sleep, 1995. **18**(2): p. 115-26.
106. Shochat, T., et al., *The SleepStrip: an apnoea screener for the early detection of sleep apnoea syndrome.* Eur Respir J, 2002. **19**(1): p. 121-6.
107. *Practice parameters for the use of portable recording in the assessment of obstructive sleep apnea. Standards of Practice Committee of the American Sleep Disorders Association.* Sleep, 1994. **17**(4): p. 372-7.

108. Chesson, A.L., Jr., et al., *The indications for polysomnography and related procedures.* Sleep, 1997. **20**(6): p. 423-87.
109. Ross, S.D., et al., *Systematic review and meta-analysis of the literature regarding the diagnosis of sleep apnea.* Sleep, 2000. **23**(4): p. 519-32.
110. *Decision Memo for Continuous Positive Airway Pressure (CPAP) Therapy for Obstructive Sleep Apnea (OSA) (CAG-00093R).* 2005; Available from: http://www.cms.hhs.gov/mcd/viewdecisionmemo.asp?id=110
111. *Portable monitoring in the diagnosis of obstructive sleep apnea.* J Clin Sleep Med, 2006. **2**(3): p. 274.
112. Trikalinos, T., et al., *Home diagnosis of obstructive sleep apnea-hypopnea syndrome*, A.f.H.R.a. Quality, Editor. 2007, AHRQ: Rockville.
113. Lux, L., B. Boehlecke, and K.N. Lohr, *Effectiveness of portable monitoring devices for diagnosing obstructive sleep apnea: Update of a systematic review.*, Agency for Healthcare Research and Quality, Editor. 2004, AHRQ: Rockville, MD.
114. Collop, N.A., et al., *Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine.* J Clin Sleep Med, 2007. **3**(7): p. 737-47.
115. Ghegan, M.D., et al., *Laboratory versus portable sleep studies: a meta-analysis.* Laryngoscope, 2006. **116**(6): p. 859-64.
116. Joint Nordic Project, *Obstructive Sleep Apnoea Syndrome. A Systematic Literature Review.* 2007: Norrkoping, Sweden.
117. Thurnheer, R., et al., *Respiratory polygraphy in sleep apnoea diagnosis. Report of the Swiss respiratory polygraphy registry and systematic review of the literature.* Swiss Med Wkly, 2007. **137**(5-6): p. 97-102.
118. Tice, J.A., *Portable devices used for home testing in obstructive sleep apnea*. 2009: San Francisco.
119. Tregear, S., M. Tiller, and J. Fontarrosa, *Obstructive sleep apnea and commercial motor vehicle driver safety*, Federal Motor Carrier Safety Administration, Editor. 2007.
120. Berry, R.B., et al., *Portable monitoring and autotitration versus polysomnography for the diagnosis and treatment of sleep apnea.* Sleep, 2008. **31**(10): p. 1423-31.
121. Skomro, R.P., et al., *Outcomes of home-based diagnosis and treatment of obstructive sleep apnea.* Chest, 2010. **138**(2): p. 257-63.
122. Whitelaw, W.A., R.F. Brant, and W.W. Flemons, *Clinical usefulness of home oximetry compared with polysomnography for assessment of sleep apnea.* Am J Respir Crit Care Med, 2005. **171**(2): p. 188-93.
123. Mulgrew, A.T., et al., *Diagnosis and initial management of obstructive sleep apnea without polysomnography: a randomized validation study.* Ann Intern Med, 2007. **146**(3): p. 157-66.
124. Watkins, M.R., et al., *Correlation between screening for obstructive sleep apnea using a portable device versus polysomnography testing in a commercial driving population.* J Occup Environ Med, 2009. **51**(10): p. 1145-50.
125. Pittman, S.D., 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): p. 923-33.
126. Golpe, R., A. Jimenez, and R. Carpizo, *Home sleep studies in the assessment of sleep apnea/hypopnea syndrome.* Chest, 2002. **122**(4): p. 1156-61.
127. Baltzan, M.A., et al., *Accuracy of oximetry with thermistor (OxiFlow) for diagnosis of obstructive sleep apnea and hypopnea.* Sleep, 2000. **23**(1): p. 61-9.

128. Ryan, P.J., 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): p. 972-5.
129. Series, F., et al., *Utility of nocturnal home oximetry for case finding in patients with suspected sleep apnea hypopnea syndrome.* Ann Intern Med, 1993. **119**(6): p. 449-53.
130. *Executive summary on the systematic review and practice parameters for portable monitoring in the investigation of suspected sleep apnea in adults.* Am J Respir Crit Care Med, 2004. **169**(10): p. 1160-3.
131. Blackman, A., C. McGregor, and R. Dales, *Canadian Sleep Society/Canadian Thoracic Society position paper on the use of portable monitoring for the diagnosis of obstructive sleep apnea/hypopnea in adults.* Can Respir J, 2010. **17**(5): p. 229-32.
132. Chesson, A.L., Jr., R.B. Berry, and A. Pack, *Practice parameters for the use of portable monitoring devices in the investigation of suspected obstructive sleep apnea in adults.* Sleep, 2003. **26**(7): p. 907-13.
133. Collop, N.A., *Portable monitoring for the diagnosis of obstructive sleep apnea.* Curr Opin Pulm Med, 2008. **14**(6): p. 525-9.
134. Kuna, S.T., *Portable-monitor testing: an alternative strategy for managing patients with obstructive sleep apnea.* Respir Care, 2010. **55**(9): p. 1196-215.
135. Marcos, J.V., et al., *Automated detection of obstructive sleep apnoea syndrome from oxygen saturation recordings using linear discriminant analysis.* Med Biol Eng Comput, 2010. **48**(9): p. 895-902.
136. Marcos, J.V., et al., *Applying neural network classifiers in the diagnosis of the obstructive sleep apnea syndrome from nocturnal pulse oximetric recordings.* Conf Proc IEEE Eng Med Biol Soc, 2007. **2007**: p. 5174-7.
137. Marcos, J.V., et al., *Radial basis function classifiers to help in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry.* Med Biol Eng Comput, 2008. **46**(4): p. 323-32.
138. Marcos, J.V., et al., *A classification algorithm based on spectral features from nocturnal oximetry and support vector machines to assist in the diagnosis of obstructive sleep apnea.* Conf Proc IEEE Eng Med Biol Soc, 2009. **2009**: p. 5547-50.
139. Marcos, J.V., et al., *Assessment of four statistical pattern recognition techniques to assist in obstructive sleep apnoea diagnosis from nocturnal oximetry.* Med Eng Phys, 2009. **31**(8): p. 971-8.
140. Marcos, J.V., et al., *Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry.* Comput Methods Programs Biomed, 2008. **92**(1): p. 79-89.
141. Marcos, J.V., et al., *The classification of oximetry signals using Bayesian neural networks to assist in the detection of obstructive sleep apnoea syndrome.* Physiol Meas, 2010. **31**(3): p. 375-94.
142. Mendez, M.O., et al., *Sleep apnea screening by autoregressive models from a single ECG lead.* IEEE Trans Biomed Eng, 2009. **56**(12): p. 2838-50.
143. Mendez, M.O., et al., *Detection of sleep apnea from surface ECG based on features extracted by an autoregressive model.* Conf Proc IEEE Eng Med Biol Soc, 2007. **2007**: p. 6106-9.
144. Morillo, D.S., et al., *Poincare analysis of an overnight arterial oxygen saturation signal applied to the diagnosis of sleep apnea hypopnea syndrome.* Physiol Meas, 2009. **30**(4): p. 405-20.
145. Ndegwa, S., M. Clark, and C. Argaez, *Portable monitoring devices for diagnosis of obstructive sleep apnea at home: Review of accuaracy, cost-effectiveness, guidelines, and coverage in Canada.*, C.A.f.D.a.T.i. Health, Editor. 2009: Ottawa.

146. Polese, J.F., et al., *Portable monitoring devices in the diagnosis of obstructive sleep apnea: current status, advantages, and limitations.* J Bras Pneumol, 2010. **36**(4): p. 498-505.
147. Poupard, L., et al., *Use of high-frequency peak in spectral analysis of heart rate increment to improve screening of obstructive sleep apnoea.* Sleep Breath, 2010.
148. Ramachandran, S.K. and L.A. Josephs, *A meta-analysis of clinical screening tests for obstructive sleep apnea.* Anesthesiology, 2009. **110**(4): p. 928-39.
149. Roche, F., et al., *Analysis of the interbeat interval increment to detect obstructive sleep apnoea/hypopnoea.* Eur Respir J, 2007. **29**(6): p. 1206-11.
150. Salisbury, J.I. and Y. Sun, *Rapid screening test for sleep apnea using a nonlinear and nonstationary signal processing technique.* Med Eng Phys, 2007. **29**(3): p. 336-43.
151. Stein, P.K., et al., *A simple method to identify sleep apnea using Holter recordings.* J Cardiovasc Electrophysiol, 2003. **14**(5): p. 467-73.
152. Sun, J., et al., *Identification of obstructive sleep apnea syndrome by ambulatory electrocardiography: clinical evaluation of time-domain and frequency-domain analyses of heart rate variability in Chinese patients.* Cell Biochem Biophys, 2011. **59**(3): p. 165-70.
153. Sunwoo, B. and S.T. Kuna, *Ambulatory management of patients with sleep apnea: is there a place for portable monitor testing?* Clin Chest Med, 2010. **31**(2): p. 299-308.
154. Suzuki, T., et al., *Development of a sleep apnea event detection method using photoplethysmography.* Conf Proc IEEE Eng Med Biol Soc, 2010. **2010**: p. 5258-61.
155. Tsai, W.H., et al., *A decision rule for diagnostic testing in obstructive sleep apnea.* Am J Respir Crit Care Med, 2003. **167**(10): p. 1427-32.

Appendix A: Search Summaries

Appendix B: Retrieval Criteria

Retrieval Criteria for Key Question #1

- Article must have been published in the English language
- Article must have enrolled 10 or more subjects
- Article must describe sleep studies that were performed with both facility-based PSG and algorithms/models designed to predict the severity of OSA in the same patients, either simultaneously or within 3 months of first measurement.

Retrieval Criteria for Key Question #2

- Article must have been published in the English language.
- Article must have enrolled 10 or more subjects.
- Article must be a systematic review of the literature or randomized controlled trial.
- Systematic reviews must include sleep studies that were performed with both facility-based PSG and portable monitors in the same patients, either simultaneously or within 3 months of the first measurement.
- RCTs must compare portable monitoring with facility-based PSG for the diagnosis and treatment of OSA with CPAP.

Appendix C: Inclusion Criteria

Inclusion Criteria for Key Question #1

- Article must have been published in the English language.
- Article must be a full-length article. Abstracts and letters to the editor will not meet this inclusion criterion.
- Article must have enrolled 10 or more subjects.
- Article must have enrolled subjects aged ≥18.
- Article must have enrolled individuals with obstructive sleep apnea only; no central apneas.
- Article must describe sleep studies that were performed with both facility-based PSG and algorithms/models designed to predict the severity of OSA in the same patients, either simultaneously or within 3 months of the first measurement.
- Article must have presented the actual algorithm/model, not merely referred to it
- Article must describe an algorithm/model consisting of variables that are conveniently measurable in a regular clinic setting. Algorithms/models with variables requiring specialized training and/or expensive equipment (≥ $300) to measure were excluded.
- Article must report outcome in terms of sensitivity and specificity of the algorithm/model relative to PSG AHI or RDI, or present data in a manner that allows one to calculate sensitivity and specificity of the algorithm/model.

Inclusion Criteria for Key Question #2

- Article must have been published in the English language. Moher et al.(252) have demonstrated that exclusion of non-English language studies from meta-analyses has little impact on the conclusions drawn. Juni et al.(251) found that non-English studies typically were of lower methodological quality and that excluding them had little effect on effect size estimates in the majority of meta-analyses they examined. Although we recognize that in some situations exclusion of non-English studies could lead to bias, we believe that the few instances in which this may occur do not justify the time and cost typically necessary for translation of studies to identify those of acceptable quality for inclusion in our reviews.(251,252)
- Article must be a full-length article. Abstracts and letters to the editor will not meet this inclusion criterion.
- Article must have enrolled 10 or more subjects.
- Article must have enrolled subjects aged ≥18.
- Individuals with obstructive sleep apnea only, no central apneas.
- Studies that evaluated both obstructive sleep apnea and other sleep disordered individuals were included as long as data for obstructive sleep apnea subjects could be analyzed separately from that of other subject populations.
- Article must be a systematic review of the literature or randomized controlled trial.

- Systematic reviews must include sleep studies that were performed with both facility-based PSG and portable monitors in the same patients, either simultaneously or within 3 months of the first measurement.
- RCTs must compare portable monitoring with facility-based PSG for the diagnosis and treatment of OSA with CPAP.

Appendix D: Excluded Articles

Table D-1. Excluded Articles for Key Question #1

Reference		Reasons for Exclusion
Bliwise et al.	1991	Irrelevant
Brown et al.	2010	No algorithm equation presented
Caffo et al.	2010	No algorithm equation presented
de Silva et al.	2011	Variable measurement expensive and/or impractical
Campos-Rodriguez	2008	Commentary
Deegan & McNicholas	1996	Diagnostic performance measures not reported
Drager et al.	2010	Diagnostic performance measures not reported
Flemons & McNicholas	1997	Background
Flemons et al.	1993	Diagnostic performance measures not reported
Friedman et al.	2010	Irrelevant
Furukawa et al.	2009	Irrelevant
Haponik et al.	1984	Variable measurement expensive and/or impractical
Herzog et al.	2009	Variable measurement expensive and/or impractical
Hoffstein & Szalai	1993	Diagnostic performance measures not reported
Hsu et al.	2005	Variable measurement expensive and/or impractical
Hukins	2010	No algorithm equation presented
Johnson et al.	2011	Variable measurement expensive and/or impractical
Katz et al.	1990	Variable measurement expensive and/or impractical
Kolotkin et al.	2011	No algorithm equation presented
Kripke et al.	2010	Diagnostic performance measures not reported
Lam et al.	2005	Diagnostic performance measures not reported
Lee et al.	2009	Variable measurement expensive and/or impractical
Len et al.	2006	Diagnostic performance measures not reported
Leng et al.	2003	Irrelevant
Maislin et al.	1995	Diagnostic performance measures not reported
Marcos et al.	2009	Variable measurement expensive and/or impractical
Martinez Garcia et al.	2003	Used portable monitor as reference standard
Mulgrew et al.	2007	Diagnostic performance measures not reported
Nuckton et al.	2006	No algorithm equation presented
Ogretmenoglu et al.	2005	No algorithm equation presented
Parks et al.	2009	Diagnostic performance measures not reported

Reference		Reasons for Exclusion
Penzel et al.	2002	Diagnostic performance measures not reported
Polat et al.	2008	Variable measurement expensive and/or impractical
Polat, Yosunkaya & Güneş	2008	Variable measurement expensive and/or impractical
Ramachandran et al.	2010	No algorithm equation presented
Rao et al.	2006	Background
Ray et al.	2010	Variable measurement expensive and/or impractical
Rodrigues et al.	2007	Use of special study population (acromegaly)
Romero et al.	2010	No algorithm equation presented
Rollheim et al.	1997	Background
Rouatbi et al.	2009	Diagnostic performance measures not reported
Saarelainen et al.	2003	Variable measurement expensive and/or impractical
Salisbury & Sun	2007	Variable measurement expensive and/or impractical
Santaolalla Montoya et al.	2007	Variable measurement expensive and/or impractical
Scharf et al.	1990	Irrelevant
Schafer et al.	1997	Variable measurement expensive and/or impractical
Senny et al.	2008	Variable measurement expensive and/or impractical
Serafini et al.	2000	Diagnostic performance measures not reported
Shizuku et al.	2008	No algorithm equation presented
Soriano et al.	2010	Background
Stoohs et al.	1996	Background
Sun et al.	2010	No algorithm equation presented
Takegami et al.	2009	Used oximetry as reference standard
Talmage et al.	2008	No algorithm equation presented
Tami et al.	1998	Diagnostic performance measures not reported
Ten Have & Bixler	1997	Irrelevant
Torre-Bouscoulet et al.	2009	Irrelevant
Ward et al.	2006	Diagnostic performance measures not reported
Weiss et al.	2005	No distinction between levels of OSA severity
Whitney et al.	1998	Irrelevant
Xie et al.	2011	No algorithm equation presented
Yao et al.	2006	No algorithm equation presented
Young & McDonald	2004	Diagnostic performance measures not reported

Table D-2. Excluded Articles for Key Question #2

Study	Year	Reason for exclusion
AASM [130]	2004	Duplicate review
Blackman et al. [131]	2010	Not systematic review/RCT
Chesson et al. [132]	2003	Duplicate review
Collop et al. [133]	2008	Not systematic review/RCT
Hornero et al.	2007	Not systematic review/RCT
Kuna et al. [134]	2010	Not systematic review/RCT
Marcos et al. [135]	2010	Not systematic review/RCT
Marcos et al. [136]	2007	Not systematic review/RCT
Marcos et al. [137]	2008	Not systematic review/RCT
Marcos et al. [138]	2009	Not systematic review/RCT
Marcos et al. [139]	2009	Not systematic review/RCT
Marcos et al. [140]	2008	Not systematic review/RCT
Marcos et al. [141]	2010	Not systematic review/RCT
Mendez et al. [142]	2009	Not systematic review/RCT
Mendez et al. [143]	2007	Not systematic review/RCT
Morillo et al. [144]	2009	Not systematic review/RCT
Ndegwa et al. [145]	2009	Not systematic review/RCT (Review of reviews)
Parks et al. [35]	2009	Not systematic review/RCT
Polese et al. [146]	2010	Not systematic review/RCT
Poupard et al. [147]	2010	Not systematic review/RCT
Ramachandran et al. [148]	2009	Not portable monitoring
Roche et al. [149]	2007	Not systematic review/RCT
Ross et al. [109]	2000	Updated in another review
Salisbury et al. [150]	2007	Not systematic review/RCT
Stein et al. [151]	2003	Not systematic review/RCT
Sun et al. [152]	2011	Not systematic review/RCT
Sunwoo et al. [153]	2010	Not systematic review/RCT
Suzuki et al. [154]	2010	Not systematic review/RCT
Talmage et al. [97]	2008	Not systematic review/RCT
Tsai et al. [155]	2003	Not systematic review/RCT