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

W. Ward Flemons, MD; Michael R. Littner, MD, FCCP; James A. Rowley, MD, FCCP; Peter Gay, MD, FCCP; W. McDowell Anderson, MD, FCCP; David W. Hudgel, MD, FCCP; R. Douglas McEvoy, MBBS, MD; and Daniel I. Loube, MD, FCCP

(CHEST 2003; 124:1543–1579)

Key words: diagnosis; likelihood ratios; polysomnography/methods; research design; review literature; sensitivity and specificity; sleep apnea syndromes

Abbreviations: AASM = American Academy of Sleep Medicine; ACCP = American College of Chest Physicians; AHI = apnea-hypopnea index; AHRQ = Agency for Healthcare Research and Quality; ATS = American Thoracic Society; CPAP = continuous positive airway pressure; EMG = electromyogram; ERC = Evidence Review Committee; LR = likelihood ratio; RDI = respiratory disturbance index; REM = rapid eye movement; ROC = receiver operating characteristic; RTI-UNC = Research Triangle Institute-University of North Carolina

1.0 Introduction/Background

Sleep apnea is a common disorder that affects both children and adults. It is characterized by periods of breathing cessation (apnea) and periods of reduced breathing (hypopnea). Both types of events have similar pathophysiology and are generally considered to be equal with respect to their impact on patients.[1] The most common form of sleep apnea, called *obstructive sleep apnea*, is caused by the partial or complete collapse of the upper airway. There are several methods of quantifying the severity of the disorder such as measuring the number of apneas and hypopneas per hour of sleep (*ie*, the apnea-hypopnea index [AHI]), the severity of oxygen desaturation during sleep, or the severity of the most commonly associated symptom, daytime somnolence. The prevalence of an AHI of ≥ 5 was 24% in men and 9% in women aged 30 to 60 years in the Wisconsin Sleep Cohort Study.[2] The prevalence of symptomatic sleep apnea (*ie*, AHI of ≥ 5 with excessive daytime somnolence) for men and women was 4% and 2%, respectively.[2] The standard approach to diagnosis is in-laboratory, technician-attended polysomnography that monitors, at a minimum, sleep time and respiration. Polysomnography requires technical expertise, and is labor-intensive and time-consuming. Timely access is a problem for many patients, the majority of whom continue to have undiagnosed sleep apnea. In the Wisconsin sleep cohort study,[3] 93% of women and 82% of men with moderate-to-severe sleep apnea did not receive diagnoses. Thus, there is a growing interest in alternative approaches to diagnosis, such as portable monitoring, that have been proposed as a substitute for polysomnography in the diagnostic assessment of

For related article see page 1535

*From the Faculty of Medicine (Dr. Flemons), University of Calgary, Calgary, AB, Canada; David Geffen School of Medicine (Dr. Littner), University of California Los Angeles, Los Angeles, CA; Division of Pulmonary, Critical Care, and Sleep Medicine (Dr. Rowley), Department of Medicine, Wayne State University School of Medicine, Detroit, MI; Pulmonary, Critical Care, and Sleep Medicine (Dr. Gay), Mayo Clinic, Rochester, MN; the University of South Florida College of Medicine (Dr. Anderson), Tampa, FL; Case Western Reserve University (Dr. Hudgel), Cleveland, OH; Adelaide Institute for Sleep Health (Dr. McEvoy), Repatriation General Hospital, Daw Park, Australia; and the Swedish Medical Center (Dr. Loube), Seattle, WA.

Supplemental evidence tables are available online at http://www.chestjournal.org/cgi/content/full/124/4/1543/DC1.

Manuscript received May 1, 2003; revision accepted May 2, 2003.

Correspondence to: W. Ward Flemons, MD, Faculty of Medicine, University of Calgary, 1403 Twenty-Ninth St NW, Calgary, AB, Canada T2N 2T9; e-mail: flemons@ucalgary.ca

patients with suspected sleep apnea. The term *portable monitoring* encompasses a wide range of devices that can record as many signals as does attended polysomnography or only one signal, such as with oximetry (see section 1.1). When EEG and electromyogram (EMG) signals are recorded, sleep staging can be performed that provides a denominator for the AHI. More commonly, EEG and EMG signals are not recorded by portable monitors, in which case breathing events are usually quantified per hour of monitoring time as a respiratory disturbance index (RDI). The use of portable monitoring to assess patients suspected of having sleep apnea is controversial and has been the subject of previous reviews of the literature.[4–8] Since the last review was completed, there have been additional research studies published and more standardized methods developed for rating the evidence of studies on diagnostic tests.

The American Thoracic Society (ATS), the American College of Chest Physicians (ACCP), and the American Academy of Sleep Medicine (AASM) individually planned to review and update the evidence on the diagnostic validity of portable monitors for diagnosing sleep apnea in adults. At a conference hosted by the ACCP in September 2000, an initial proposal to collaborate on this project was discussed by all three organizations that eventually led to a formal agreement to cosponsor a working group and to hire an evidence-based practice center to produce a detailed literature search and evidence review on the use of portable monitors for investigating patients with suspected sleep apnea. Two other organizations, the National Association for the Medical Direction of Respiratory Care and the Australasian Sleep Association, agreed to participate as liaison organizations and appointed members to the committee structure. Detailed conflict-of-interest guidelines were established that prevented anyone with a link to industries that made commercially available sleep apnea portable monitors from working on this project (details available on request). The ACCP accepted administrative responsibility for the working group. The following three committees were created with at least one representative from each sponsoring organization: (1) Steering Committee, Nancy Collop (Chair), Patrick Strollo, and John Shepard; (2) Evidence Review Committee (ERC), Ward Flemons (Chair), James Rowley, Michael Littner, William Anderson, David Hudgel, Dan Loube, Peter Gay, and Doug McEvoy; and (3) Guideline Committee, Andrew Chesson (Chair), Allan Pack, and Richard Berry.

Funding for this project, including an evidence review that was performed under contract by a team of evidence-based researchers at RTI International and the University of North Carolina at Chapel Hill (RTI-UNC), was provided completely and jointly by the ATS, the ACCP, and the AASM.

Three previous reviews of sleep apnea portable monitoring devices have been published. In 1994, the AASM (formerly the American Sleep Disorders Association) published a description of 23 studies[4] that reported some features of portable monitoring. In this review, sleep studies were categorized into the following four types: type 1, standard polysomnography; type 2, comprehensive portable polysomnography; type 3, modified portable sleep apnea testing; and type 4, continuous single-bioparameter or dual-bioparameter recording (see section 1.1.1). In 1997, the AASM published practice parameters[5] and a review[6] for indications for polysomnography and related procedures that included a section on type 3 and type 4 studies. Based on the review, the practice parameters recommended that attended type 3 studies were potentially appropriate in patients with a high pretest probability (*eg*, $> 70\%$) of sleep apnea. The parameters recommended that negative type 3 monitor studies in symptomatic patients be followed up with a full polysomnogram. The parameters did not recommend type 4 studies for the investigation of suspected sleep apnea. In 1997, the Agency for Healthcare Research and Quality (AHRQ) [formerly, the Agency for Health Care Policy and Research] in the United States commissioned a systematic review of the research on the diagnosis of sleep apnea.[7,8] Part of that review focused on studies of portable monitors (25 studies), including oximetry (12 studies), and included articles published from 1980 to November 1, 1997.[7,8] As part of this systematic review, the quality of each reviewed study was rated using a scale that the authors developed. This was a potentially helpful addition to the AASM reviews because it attempted to identify and account for biases that may undermine the validity of the findings and conclusions of a study.

Over the past decade, there has been increasing interest in developing methods to rate the quality of research studies, especially when a systematic review is undertaken. There has been more work published on the methods for rating the research evidence of therapeutics studies than the rating of diagnostic testing studies. The ACCP/ATS/AASM working group decided it was important to update the literature review from 1997 as well as to update the system used to rate the quality of the research evidence on portable monitoring. The method published by Sackett et al[9] in 2000 for rating evidence of research on diagnostic tests was used because it closely aligns with the accepted methods used for rating the quality of articles on therapeutics and prognosis. In addition, it focuses on the following key

aspects of design for studies of diagnostic tests: avoiding selection bias (by using a consecutively referred sample of patients); blinding of the interpreters; and avoidance of verification bias (by performance of the reference standard on all subjects).

An increasing amount of research has been published comparing some type of portable monitoring for sleep apnea with polysomnography. From 1990 to 2001, a total of 51 articles that met preselected inclusion/exclusion criteria for being included in this latest systematic review of portable monitoring for sleep apnea have been published in the English literature. These articles were rated with respect to the level of evidence (*ie*, I, II, III, or IV) based, in part, on the approach published by Sackett et al[9] (see section 1.5). The majority of studies (30 of 51 studies) were of higher quality (*ie*, levels I and II), but there is not yet a trend of this percentage increasing over time. In Figure 1, the number of level I studies (best quality) and level II studies, as well as the total number of studies published on portable monitoring are shown over time.

The goal of a systematic review is to summarize a body of literature to aid in reaching conclusions about a particular practice in medicine. A common approach used to synthesize evidence is meta-analysis. This approach was used in the AHRQ commissioned review, and the results were reported in the form of summary receiver operating characteristic (ROC) curves.[8] The current working group decided against a meta-analysis of results because there was too much heterogeneity between studies with respect to types of signals measured (Table 1[10–60] and section 1.1.1), criteria used to define a *breathing event* (section 1.1.2), how signals from portable monitors were scored (section 1.1.3), and study quality (section 1.5). Therefore, the working group elected to summarize and report the details of each study to allow for conclusions to be drawn about the evidence without combining results across studies in a formal meta-analysis. Study data were synthesized into tables and were categorized as follows: (1) monitor type (section 1.1); (2) location of the study (unattended at home vs attended in the sleep laboratory) [section 1.4.1]; and (3) evidence level and quality rating (section 1.5).

Three primary and four secondary areas are addressed in this report. The primary areas are as follows:

1. The utility of portable monitors in reducing the probability that a patient has an abnormal AHI (rule out the disorder) [section 4.1.1];
2. The utility of portable monitors in increasing the probability that a patient has an abnormal AHI (rule in the disorder) [section 4.1.2];
3. The utility of portable monitors in both reducing and increasing the probability that a patient has an abnormal AHI (rule out and rule in the disorder) [section 4.1.3].

The secondary areas are as follows:

1. The reproducibility of portable monitor results [section 4.2.1];
2. The cost benefit of portable monitors [section 4.2.2];
3. The failure rates of portable monitors [section 4.2.3];
4. The patient populations studied and the generalizability of findings [section 4.2.4].

Finally, it was the goal of this working group to outline the deficiencies in the current evidence on portable monitors for the investigation of patients with suspected sleep apnea, to describe opportunities for future research, and to highlight key methodological issues that should be addressed by future researchers, journal editors, reviewers, and readers of this literature.

1.1. Portable Monitoring

1.1.1 Types of monitors

Portable monitors were classified according to the approach used in the 1994 American Sleep Disorders Association review.[4] Type 1 (standard polysomnography) was considered the reference standard to which the other monitor types were compared. The

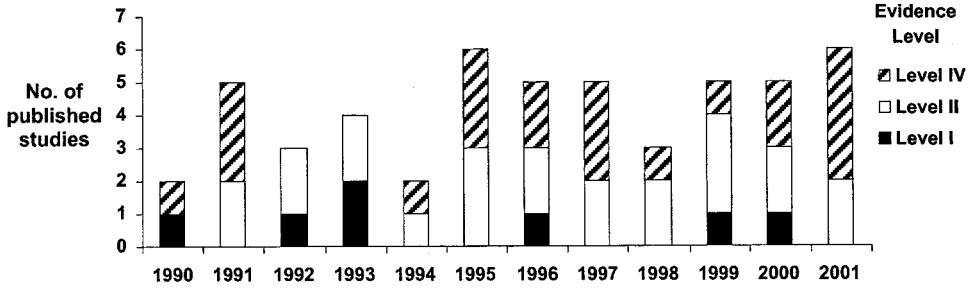

FIGURE 1. Quality of published studies on portable monitoring for sleep apnea, 1990 to 2001.

Table 1—*Breathing Event Definitions and Study Information**

Scoring Channel and Criteria					Oxim Samp Freq, Hz	PSG Event Definition	Evid Level	Quality Rating	Study/Year	Male, %	Mean AHI	Mean BMI	PM Type	PM Name†	Score Method	Ref. No.
Primary Criteria			Secondary Criteria													
Flow (NP)	↓ > 50%					Complete ↓ flow (therm) or ↓ thor/abd RIP > 50%	II	a	Rees et al/1998 (n = 20)	100	39	31	4	AutoSet	Auto	10
						↓ Flow (therm) > 50%	II	a	Gugger et al/1997 (n = 67)	87	26	31	4	AutoSet	Auto	11
						↓ Flow (PT) + ↓ O_2 > 3% or arousal	II	a	Mayer et al/1998 (n = 95)	83	43	31	4	AutoSet	Auto	12
						↓ Flow (therm) > 50%	II	b	Kiely et al/1996 (n = 36)	75	19.4	28	4	AutoSet	Auto	13
						NR	IV	b	Bradley et al/1995 (n = 31)	84	25	30	4	AutoSet	Auto	14
	↓ > 80%	or	Flow + ↓ O_2	50% + > 3%	NR	↓ Flow (NS) > 80% or ↓ flow > 50% + ↓ O_2 > 3% or arousal	IV	a	Bagnato et al/2000 (n = 56)	80	38	31.3	4	AutoSet	Auto	15
	No flow					NR	IV	c	Gugger et al/1995 (n = 27)	NR	NR	29	4	AutoSet	Auto	16
		or	Flow (NP)	NS		Complete ↓ flow (PT) or ↓ flow—not specified	IV	d	Fleury et al/1996 (n = 44)	77	NR	29	4	AutoSet	Auto	17
Flow (therm)	Discernible ↓	and	↓ O_2	> 1%	NR	Discern ↓ flow (therm) + ↓ O_2 > 3%	II	a	Redline et al/1991 (n = 25)	83	NR	31	3	Edentec	Manual	18
				> 2%	NR	Complete ↓ flow (therm) or discern ↓ flow + ↓ O_2 > 3% or arousal	IV	a	Fry et al/1998 (n = 93)	64	29	NR	2	DigiTrace	Manual	19
	↓ > 25%					↓ Flow (therm) > 25%	IV	b	Portier et al/2000 (n = 78)	82	26	31	2	MiniSomno	Manual	20
	↓ > 40%					↓ Flow (therm) > 40%	I	a	Zucconi et al/1996 (n = 29)	69	32	31	3	Microdigitrap	Comb	21
	↓ > 50%					↓ Flow (therm or $ETCO_2$) > 50%	I	a	Emsellem et al/1990 (n = 63)	NR	NR	NR	3	EdenTrace	Manual	22
						↓ Flow (therm) > 50%	II	a	Man and Kang/1995 (n = 104)	80	NR	NR	3	Poly G	Manual	23
						↓ Flow (therm) > 50%	IV	a	Ancoli-Israel et al/1997 (n = 34)	89	42	NR	3	Night Watch	Auto + manual	24
		or	RIP	↓ > 50%	1	Complete ↓ flow (therm) or ↓ flow > 50% or ↓ thor/abd RIP > 50%	II	a	Mykytyn et al/1999 (n = 20)	100	NR	32	2	Compumedics	Manual	25
	No flow	or	Flow + ↓ O_2	Discernible + > 1%	NR	Complete ↓ flow (therm) or discern ↓ flow or paradox thor/abd + ↓ O_2 (NS) or arousal	II	a	Parra et al/1997 (n = 89)	82	34	29	3	EdenTrace	Manual	26
				Discernible + > 2%	NR	Complete ↓ flow (therm) or discern ↓ flow + ↓ O_2 > 2% or arousal	II	b	Ballester et al/2000 (n = 116)	56	10	26	3	Sibel-home monitor	Manual	27
				Discernible + > 3%	NR	Complete ↓ flow (therm) or ↓ flow > 50% + ↓ O_2 > 3%	II	a	Ficker et al/2001 (n = 51)	86	24	29	3	Somnocheck	Manual	28
Flow (therm)	No flow	or	Flow + ↓ O_2	> 3% and > 1%	4	Complete ↓ flow (therm) or ↓ flow > 50% + ↓ O_2 > 3%/1% or arousal	II	a	White et al/1995 (n = 100)		28/31	31/33	3	Night Watch	Auto + manual	29
			Chest wall	↓ > 50%		Complete ↓ flow (therm) or ↓ thor/abd RIP > 50%	IV	b	Whittle et al/1997 (n = 78)	93	NR	31	3	EdenTrace	Manual	30
	NR					↓ Flow (NS) > 50%	I	a	Verse et al/1992 (n = 53)	92	18	27	3	Poly-MESAM	Auto + manual	31
						NR	IV	d	Orr et al/1994 (n = 40)	NR	26	NR	2	Sleep I/T	Auto	32

Table 1—*Continued

Scoring Channel and Criteria						Oxim Samp Freq, Hz	PSG Event Definition	Evid Level	Quality Rating	Study/Year	Male, %	Mean AHI	Mean BMI	PM Type	PM Name†	Score Method	Ref. No.
Primary Criteria				Secondary Criteria													
Flow (therm) + ↓ O_2	↓ 50% + > 3% ↓					10	↓ Flow (therm) > 90% or ↓ flow > 50% + ↓ O_2 > 3%	II	a	Baltzan et al/2000 (n = 97)	75	45	28.4	4	Oxiflow	Comb	33
O_2 saturation	> 2% ↓					NR	NR	II	a	Stoohs and Guilleminault/ 1992 (n = 56)	82	21	NR	4	Mesam IV	Auto	34
						NR	Complete ↓ flow (NS) or ↓ flow > 50% + ↓ O_2 > 3% or arousal	II	d	Gurubhagavatula et al/2001 (n = 359)	68	26	32	4	NR	Manual	35
			and	Resaturation	> 2%	1	Complete ↓ flow (therm) or ↓ flow > 50% + ↓ O_2 > 3% or arousal	I	a	Vazquez et al/2000 (n = 241)	78	26	32	4	SnoreSat	Auto	36
						2	NR	II	b	Yamashiro et al/ 1995 (n = 269)	70	NR	NR	4	Biox 3740	Manual	37
						0.5 and 0.08	↓ O_2 > 3%	IV	a	Wiltshire et al/2001 (n = 100)	36	NR	NR	4	Biox 3740	Auto	38
	> 3% ↓			O_2 saturation	> 1% and HR Δ	0.5	Complete ↓ flow (NS) or ↓ thor/abd > 30% + ↓ O_2 > 3% or arousal	IV	b	Schafer et al/1997 (n = 114)	88	29	31	4	Mesam IV	Manual	39
					CT < 90% > 1%	0.08	↓ Flow (NS) > 50%	I	d	Gyulay et al/1993 (n = 98)	79	NR	30	4	Biox 3700	Auto or manual	40
			or		CT < 90% > 0.8%	0.2	Complete ↓ flow (therm) or discern ↓ flow + ↓ O_2 > 3% or arousal	II	c	Golpe et al/1999 (n = 116)	90	24	30	4	Minolta Pulseox 7	Auto + manual	41
				O_2 saturation	< 90%	0.16	Complete ↓ flow (NS)	IV	c	Williams et al/1991 (n = 36)	NR	40	NR	4	Biox 3700 IV/ Nellcor N100	Manual	42
				Snoring pause or HR variability	10–120	NR	Complete ↓ flow (therm) or discern ↓ flow + ↓ O_2 > 3% or arousal	II	a	Esnaola et al/1996 (n = 150)	89	43	30	4	Mesam IV	Auto or manual	43
					> 10% HR Δ	NR	Complete ↓ flow (NS) or ↓ thor/abd > 50%	II	a	Koziej et al/1994 (n = 56)	91	NR	NR	4	Mesam IV	Auto or manual	44
	> 4% ↓					0.16	Complete ↓ flow (NS) or ↓ flow or thor/abd > 50% + ↓ O_2 > 3% or arousal	I	b	Chiner et al/1999 (n = 275)	90	42	32	4	Nellcor N200	Comb	45
						NR	NR	IV	d	Bonsignore et al/ 1990 (n = 73)	74	NR	NR	4	NR	NR	46
			or	Resaturation	> 2%	1	Complete ↓ flow (therm) or ↓ thor/abd > 50% + ↓ O_2 > 2%	IV	c	Rauscher et al/1991 (n = 53)	NR	NR	NR	4	Mesam with Minolta Pulsox 7	Auto	47
	> 5% ↓					NR	NR	II	c	Cooper et al/1991 (n = 41)	61	NR	NR	4	Biox IIa	Manual	48
	Delta index	> 0.6				0.08	Complete ↓ flow (NS) or ↓ flow > 50% + ↓ O_2 (NS) or arousal	IV	c	Levy et al/1996 (n = 301)	NR	NR	32	4	Biox 3700	Auto	49
		> 0.8				0.08	NR	IV	d	Pepin et al/1991 (n = 26)	NR	51	NR	4	Biox 3700	Auto	50
		> 0.4	or	CT < 90%	> 1%	0.08	NR	IV	d	Olson et al/1999 (n = 906)	NR	40	NR	4	Biox 4700	Auto	51

Table 1—*Continued

Scoring Channel and Criteria					Oxim Samp Freq, Hz	PSG Event Definition	Evid Level	Quality Rating	Study/Year	Male, %	Mean AHI	Mean BMI	PM Type	PM Name†	Score Method	Ref. No.
Primary Criteria			Secondary Criteria													
O_2 saturation	PA threshold 0.7%	or	O_2 saturation	> 3% ↓	0.2	Complete ↓ flow (NS) or ↓ flow > 50% + ↓ O_2 > 3%	II	a	Zamarron et al/1999 (n = 233)	80	40	30	4	Criticare 504	Auto	52
	No criteria				2	Complete ↓ flow (NS) or ↓ thor/abd > 50% + ↓ O_2 > 4%	I	b	Series et al/1993 (n = 240)	90	38	32	4	Biox IVA	Manual	53
					NR	NR	II	a	Douglas et al/1992 (n = 200)	80	NR	NR	4	Ohmeda 3700	Auto	54
					NR	Complete ↓ flow (NS) or ↓ thor/abd > 50% + ↓ O_2 > 2%	II	b	Rauscher et al/1993 (n = 116)	82	NR	NR	4	Minolta Pulsox 7	Auto	55
					1	Complete ↓ flow (therm) + ↓ O_2 > 4% or ↓ flow (therm) > 25% + ↓ thor > 25% + ↓ abd > 15% + thor/abd paradox	IV	c	Ryan et al/1995 (n = 69)	83	NR	30	4	Minolta Pulsox 7	Manual	56
Snoring	Present	and	O_2 saturation	> 3% ↓	1	↓ thor/abd > 50% + ↓ O_2 > 3%	II	a	Issa et al/1993 (n = 29)	78	NR	31	4	SnoreSat	Auto	57
Heart rate spectral analysis	Peak power in FFT in 30–70 s window				0.2	Complete ↓ flow (NS) or ↓ flow (NS) + ↓ O_2 > 3%	IV	b	Zamarron et al/2001 (n = 197)	82	40	31	4	Criticare 504	Auto	58
Pharyngoesophageal pressure	NR					NR	IV	d	Reda et al/2001 (n = 59)	81	40	32	4	NR	NR	59
NR						Complete ↓ flow (therm) or ↓ flow > 50% + ↓ O_2 > 3%	IV	d	Claman et al/2001 (n = 42)	74	26	31	3	Bedbugg	Auto	60

*Oxim Samp Freq = oximeter sample frequency; PSG = polysomnography; Evid Level = evidence level; Pub Date = publication date; BMI = body mass index; PM = portable monitor; NR = not reported; NS = not specified; Flow (NP) = airflow measured by nasal pressure; Flow (therm) = airflow measured by thermistor; Flow (PT) = airflow measured by pneumotachometer; FFT = fast Fourier transformation; thor = thorax; abd = abdomen; paradox = paradoxical movement of the thorax and abdomen; CT = cumulative O_2 saturation time; Auto = portable monitor events were detected using an automated analysis algorithm; Manual = portable monitor events were detected by human scoring; Auto + manual = portable monitor events were detected using an automated analysis algorithm edited by human scoring; Comb = portable monitor events were detected using an automated analysis algorithm with or without editing, with both results reported; PA = pulmonary artery; ↓ = decrease; $ETCO_2$ = end-tidal CO_2; RIP = respiratory inductance plethysmography.

†Manufacturer information: AutoSet (Resmed; Poway, CA); Edentec (Nellcor; Pleasanton, CA); DigiTrace (SleepMed; Peabody, MA); MiniSomno (Nellcor); Microdigitrap; EdenTrace (Nellcor Puritan Bennett [Melville] Ltd; Kanata, ON, Canada); Poly G (CNS Inc; Whippany, NJ); Night Watch (Respironics; Murraysville, PA); Compumedics (Abbotsford, Australia); Sibel (Gembloux, Belgium); Somnocheck (Weinmann; Hamburg, Germany); Poly-MESAM (MAP GmbH; Martinsreid, Germany); Sleep I/T (CNS Inc); Oxiflow (Nellcor); Mesam IV (MAP GmbH); SnoreSat (Saga Tech; Calgary, AB, Canada; Biox monitors (Ohmeda; Madison, WI); Pulsox 7 (Minolta; Diessenhofen, Switzerland); Criticare 504 (Criticare Systems, Inc; Waukesha, WI); Ohmeda 3700 (Ohmeda); Bedbugg (Sleep Solutions Inc; Palo Alto, CA).

physiologic signals that were recorded and used to define a *breathing event* on a portable monitor varied among studies and across monitor types (Table 1[10–60]). As detailed below, type 2 monitors incorporate sleep staging as well as respiratory measures, type 3 monitors use at least three respiratory channels, and type 4 monitors use at least one respiratory channel, usually either oxygen saturation or airflow.

1.1.1.1. Type 2: comprehensive portable polysomnography

These monitors incorporate a minimum of seven channels, including EEG, electrooculogram, chin EMG, ECG or heart rate, airflow, respiratory effort, and oxygen saturation. This type of monitor allows for sleep staging and therefore for the calculation of an AHI.

1.1.1.2. Type 3: modified portable sleep apnea testing

This type of monitor incorporates a minimum of four monitored channels, including ventilation or airflow (at least two channels of respiratory movement, or respiratory movement and airflow), heart rate or ECG, and oxygen saturation.

1.1.1.3. Type 4: continuous single or dual bioparameters

Most monitors of this type measured a single parameter or two parameters, for example, oxygen saturation or airflow. A monitor that did not meet the criteria for type 3 (*ie*, a monitor that measured one to three channels or did not include airflow despite having four channels) was classified as type 4.

1.1.2. Signals used for detecting events

A challenge for the working group was to look for similarities and differences in the ways that different monitors record signals and how those signals were used to define a breathing event. As with polysomnography, there was heterogeneity with respect to defining an *abnormal breathing event* on a portable monitor (Table 1). The most common methods to detect breathing events were reduction in airflow measured by a thermistor or by a nasal pressure signal, and oxygen desaturation (several different approaches). In some circumstances, these methods were combined.

1.1.2.1. Flow

A reduction in airflow or tidal volume is the standard method for defining an apnea or hypopnea. A criterion for defining an hypopnea has been recommended to be reduction to $< 50\%$ from baseline of a valid measurement.[1] The best method for quantifying flow is a pneumotachograph. However, no portable monitors use this technology.

1.1.2.1.1. Thermistor

Thermistors sense differences in temperature and do not have a linear relationship with true airflow. Therefore, they may not be sensitive for detecting hypopneas. For these reasons, it has been recommended that for clinical research purposes thermistors not be used in polysomnography.[1] However, they are capable of sensing airflow through the nose and mouth, and it remains the most common method for defining breathing events based on a flow measurement (Table 1).

1.1.2.1.2. Nasal pressure

Nasal pressure provides a linear approximation of airflow across its complete range except at extremes. The linear relationship can be improved with a square root transformation of the signal. However, this may not be necessary if the primary use of the measure is event detection. It may not be as accurate as a thermistor in distinguishing an apnea from a hypopnea, however, in routine clinical use this distinction is not thought to be important.[1] The signal could produce false-positive events if the patient was intermittently mouth breathing, or it could be a poor quality signal if the patient was mouth breathing for long periods of time. This may require visual confirmation of apparent apneas and hypopneas, making it potentially difficult to use in an unattended study.[11–14,16,17]

1.1.2.2. Respiratory inductance plethysmography

Respiratory inductance plethysmography, when properly calibrated, can provide a measure of tidal volume. Uncalibrated, it can still be useful to detect breathing disturbances. It is used primarily during polysomnography. It was used in only one study on portable monitoring as a secondary signal (Table 1).

1.1.2.3. Oxygen saturation

Oximeters differ from other devices in important ways, particularly in the sampling frequencies and algorithms used to record oxygen saturation. Some oximeters take multiple readings, store them in memory, average them, and report a value every 3 to 12 s. Others sample and report each value at a frequency of up to 10 Hz.[33] A sampling rate of 1/12 Hz has been shown in one study to provide oxygen desaturation rates with a low number of artifacts.[38] Methods of automated analysis of the oxygen saturation signal are also variable. Most methods rely on the detection of a drop in oxygen saturation, some detect resaturation, while others use both criteria (Table 1). Some automated analyses define what baseline oxygen saturation is, but most do not. Some studies have measured the percentage of cumulative time that a patient has an oxygen saturation of

< 90% to determine whether it identifies patients with sleep apnea. Other studies have derived a "delta index" that quantifies the variability in oxygen saturation over an entire study. These last two methods do not identify specific events but instead identify patients who are likely to be experiencing apneas or hypopneas.

Oximetry analysis that is designed to detect transient drops in oxygen saturation should be more sensitive in situations in which the baseline oxygen saturation is lower because of the shape and thresholds of the oxyhemoglobin desaturation curve. Thus, patients who are studied at altitude or patients with underlying lung disease (*eg*, COPD) may show more desaturations, which could improve the sensitivity of a monitor but would likely adversely affect its specificity. Two studies have been published that evaluated COPD patients.[45,50] However, those studies did not determine how the presence of COPD affected the sensitivity and specificity of the portable monitor.

1.1.2.4. Other

One report[57] used snoring as a primary method for event detection and combined it with oxygen desaturation as a second required criterion for event detection. Other studies[43,44] have used snoring in conjunction with heart rate variability as criteria for event detection. Spectral analysis of heart rate was used in one study,[58] and a single study reported the use of pharyngoesophageal pressure measurement[59] as a method for detecting breathing events (Table 1).

1.1.3. Methods for scoring events

Studies differed in the physiologic channels monitored, the criteria used to define events, and the methods used to score events (Table 1). The majority of studies of monitors in which flow was measured by thermistor used manual scoring, while most studies of monitors in which flow was measured by nasal pressure used automated scoring. Some monitors provide automated scoring with either a computer or printed output that also allows for manual checking or editing. Some authors were explicit about how events were scored (*ie*, automated, automated with manual scoring, or only manual scoring), and in the case in which there was a component of manual scoring, by whom it was scored. However, most studies were not explicit about this. Automated scoring has the advantage that it eliminates a source of variability in results, the human recognition of events. However, polysomnograms that are used as the reference standard for defining patients with and without sleep apnea are manually scored. Therefore, an argument could be made that the automated scoring of a portable monitor is not comparable. In addition, polysomnography scoring can include an arousal from sleep as a secondary criterion (Table 1). Portable monitoring scoring, when done manually, is frequently performed in a compressed time frame from 2 to ≥ 10 min. Several studies have used automated scoring with manual editing and have reported results as various combinations of these different approaches to scoring (Table 1). Variability (interrater and intrarater) in manual scoring has not been reported (section 4.2.1). Some users may be concerned with automated scoring systems that are "black boxes," that is, they fail to identify on a record of appropriate resolution the events that were scored so that a technician or a clinician can review them, and be able to assess and edit the scoring and artifacts, and to assess the quality of the study. Details about the ability of specific monitors to display breathing events for a technician or clinician to review were not always reported in the studies that used automated scoring.

1.2. Study Location and Attendance

Portable monitors can be used in a variety of settings, including a hospital, a sleep laboratory, or in the patient's home. Portable monitors can be attended by a technician or left unattended. The role of the technician is to determine whether the portable monitor is functioning properly, to provide guidance to the patient such as encouraging patients to sleep on their backs, and for safety purposes in case there is an untoward event. With few exceptions, the research studies performed in the sleep laboratory were attended and performed simultaneously with the polysomnogram, while those performed in the home were unattended. One small study[25] performed in a sleep laboratory examined the performance of a portable type 2 monitor, attended and unattended. A second study[48] performed in a hospital was partially attended, but not simultaneously with polysomnography. One portable monitor study[29] performed at home had the opportunity for technicians to observe the study remotely and to intervene by calling the patient if there were technical problems.

1.3. Measuring Agreement

A detailed discussion of the methods for measuring agreement can be found in an accompanying publication (see page 1535). Several methods exist for evaluating the extent of agreement between two methods designed to measure the same phenomenon, including Pearson correlation coefficient, intraclass correlation coefficient, the approach of Bland and Altman of mean differences and limits of agreement, and sensitivity/specificity/likelihood ratios (LRs).

Although the Pearson correlation coefficient is widely used, it is not recommended because it is a measure of association, not agreement.[61] Intraclass correlation coefficients can be used to assess agreement,[62] however, this approach is not familiar to most clinicians and is not commonly used. The approach of Bland and Altman of calculating mean differences between two measurements is preferable to the Pearson correlation coefficient, however, the limits of agreement, which is the key descriptor that relates how well the measures agree, can also be misleading if not calculated properly. Sensitivity, specificity, and LRs have the advantage that they are in common use and are easier to understand. They address the more fundamental question of the proper classification of patients in contrast to how closely two methods agree.

Using sensitivity/specificity/LRs demands that a patient be classified as having or not having the disorder based on an arbitrary cutoff for the AHI that is variable across studies. There is a wide spectrum of the severity of breathing events at night, and the AHI captures only a single dimension. Since a substantial number of patients have indexes around the usual cutoff point, it is possible that a patient's classification might change due to expected variability in the measure (section 4.2.1). In addition, there are legitimate questions as to whether the AHI, which is derived from sleep laboratory-based polysomnography, is the correct reference standard. It is the reference standard that is most commonly used and the metric of sleep apnea severity for which there is the most published data relating to morbidity (*eg*, neurocognitive dysfunction, hypertension, and quality of life). For these reasons, it formed the basis for the systematic review that has been conducted by the ATS/AASM/ACCP working group.

The analysis of results using sensitivity, specificity, and LRs should take into account the precision of the estimates (*ie*, the calculation of confidence intervals), which are a direct reflection of sample size and study design. Studies rated with level IV evidence levels and those with small patient numbers (and wide confidence intervals) should be interpreted with caution (see Tables 3 and 4). Sensitivity, specificity, and LRs are descriptors of the operating characteristics of a test (*ie*, the degree to which the probability of disease is changed by a positive or a negative result). However, since a clinician needs to know the actual probability that the patient does or does not have a disorder (*ie*, the posttest probability or predictive value), the operating characteristics of a test have to be interpreted with the knowledge of the pretest probability (or prevalence) of the disorder (Table 2). The utility of a diagnostic test for patients with suspected sleep apnea to substitute for polysomnography can be viewed as the percentage of patients who have either a positive or negative test result and the percentage of those who have a false-positive or false-negative result, respectively (see Tables 3 to 5). Since the number of true-positive results is governed by the sensitivity and the number of false-positive results is governed by the specificity, both dictate the utility of a test, in addition to the pretest probability.

1.4. Validating Portable Monitors

Several approaches can be used to validate portable monitoring. The standard approach, and indeed what has been done to date, has been to compare portable monitoring with a reference standard, as described in the previous section. The limitation of this approach is that it assumes that sleep laboratory-based polysomnography is the optimal approach for diagnosing sleep apnea. However, this is not completely true for several reasons. From a technical perspective, patients frequently do not sleep as well in a laboratory as they do at home, and they likely spend more time on average sleeping supine. From a pragmatic perspective, the AHI correlates poorly with outcomes that are important to patients, such as quality of life and daytime sleepiness, and does not predict very well those patients who ultimately will use and thereby benefit from therapy. Therefore, a more appropriate validation study would compare the impact of portable monitoring and polysomnography on a physician's decision-making ability and outcomes important to patients. To date, there have been no studies published that have used this approach to validate the use of a portable monitor.

There are several aspects of study design and methods that, if not carefully controlled, can threaten the validity of findings and conclusions. In this review, we assigned an evidence level and quality rating for each study based on how well its design controlled possible bias (section 2.3). Other aspects of study design that affect the interpretation of findings are reviewed below.

1.4.1. Attended/nonattended monitors

The evaluation of a portable monitor in an attended setting (most often in a sleep laboratory) allows an assessment of its performance under ideal circumstances eliminating important sources of possible differences that have nothing to do with the portable monitor, such as night-to-night variability. Simultaneous assessment with polysomnography answers an important question of whether the monitor can work. If it is not also tested in an unattended setting, preferably the patient's home, the question of whether it works in the setting for which it was intended remains unanswered. When the data from a monitor used at home are compared with those

from polysomnography performed in a laboratory, the limitations of polysomnography, as a reference standard, must be kept in mind.

1.4.2. Study methodology

1.4.2.1. Describing the study population

A sufficient description of the population of patients who were studied is essential to assist readers in deciding whether the results are generalizable to their own patient population. Ideally, a broad spectrum of patients (*eg*, disease severity, age, race, men, and women) is used without the investigators participating in the selection of patients. This latter point helps to avoid selection bias. If investigators study a group of patients that they have participated in selecting (*eg*, patients referred to a sleep laboratory that the investigators refer patients to), their findings on prevalence, sensitivity, specificity, LRs, and limits of agreement could be affected and cannot be generalized to any other population of patients. There should be a clear description of who refers patients to the sleep center, the volume of referrals, as well as a description of the type of sleep center (*eg*, community-based, university setting, or Veterans Affairs hospital). If some patients are not going to be included in the study, preselected exclusion criteria should be used, with justification for those criteria. The number of patients who were referred during the time of recruitment, the number that are eligible for participation, the number who actually entered the study (with a listing of the number of patients and the reasons why patients did not participate), how many patients completed the study (with the numbers of and reasons for study drop-outs), and, finally, the percentage of cases that had uninterpretable data.

1.4.2.2. Describing portable monitor and polysomnography methods

There should be a clear description of the type of equipment used to record the signals that were used by the portable monitor and the polysomnogram. The definitions of a breathing event on polysomnography and on the portable monitor should be detailed enough to allow someone else to replicate the methods. If indexes such as the RDI are derived from scoring events, there also needs to be a clear description of these definitions. A statement about how events were scored (*eg*, automated scoring, manual scoring, epoch length, monitor type, or automated scoring with a manual review that allows for editing of results) needs to be included. Use of reference to a previous study for methods is not acceptable when the methods are part of the evaluation of the portable monitor.

1.4.2.3. Repeatability

There are many sources of variability that can limit the generalizability of a study's results. One of the most important is variability in the human recognition of events. If a monitor has an automated analysis algorithm that does not allow for manual editing, then this is not an issue. However, if there is some element of manual scoring, then the ability of two scorers and that of a single scorer repeating a review of previous scoring should be checked and reported. This intermeasurement and intrameasurement repeatability is most appropriately reported as a κ coefficient. Another source of variability that is important to study and report is night-to-night changes. This can be analyzed and reported using the approach of Bland and Altman[61] or an intraclass correlation coefficient.[62] Pearson product correlation coefficients often are used to report repeatability, but, as with reporting agreement between two different methods, it is not a recommended approach.

1.4.2.4. Avoiding bias

Several aspects of study design will dictate whether the results are more likely or less likely to be valid. There is evidence that studies of diagnostic tests with flawed designs tend to overestimate the accuracy of the test.[63] Using an appropriate series of consecutive patients controls for selection bias. Verification bias occurs when the results of one study determine whether the second study will be performed, and is avoided by ensuring that the reference standard and the diagnostic test are completed on all eligible patients. Equally important is having the diagnostic test and the reference standard interpreted and scored separately in a manner that is blinded to the results of the other test.

The *post hoc* analysis of results allows the investigators to optimize the apparent utility of the test. Performing multiple analyses reduces the reliability of statistical tests since each time an analysis is performed there is a probability by chance alone that the result will be positive (*ie*, the use of multiple analyses increases the probability of a spurious result). While there are statistical approaches to adjust for this, preselecting thresholds for a positive/negative monitoring test result prior to the study is recommended. Ideally, these thresholds have been defined in an initial study and confirmed in an independent, prospective study. Another approach is to develop thresholds using part of a patient population and to validate them on the remaining patients. Very few of the studies have adopted these approaches, and few have attempted to adjust for multiple analyses. Only two studies[30,35] validated the thresholds they used to estimate the probability of sleep apnea.

1.4.2.5. Reporting of results

The results of studies on diagnostic test accuracy should provide the number of patients who had both tests and results that clearly establish the prevalence of the disease in question, as well as the number of true-positive results, false-positive results, true-negative results, and false-negative results that will allow the calculation of sensitivity, specificity, LRs, as well as positive and negative predictive values. If different thresholds are used for the reference standard to define the presence/absence of disease, the authors should explicitly state the effect this had on prevalence as well as the operating characteristics of the diagnostic test. If multiple thresholds for a positive/negative diagnostic test result are reported, the effect of varying the threshold should be reported with a ROC curve and/or the calculation of an LR for each threshold. Finally, as with all statistical estimates, the 95% confidence intervals for the estimate (*ie*, sensitivity, specificity, and LRs) should be reported. Small patient numbers yield imprecise estimates, which are reflected by very wide confidence intervals.

1.5. Rating Levels of Evidence

The AHRQ review[7,8] on the diagnosis of sleep apnea that was released in 1999 was the first systematic review to evaluate the quality of published research in this field. The methods for rating the level of evidence of studies published on diagnostic tests have not been widely used, and the authors of the AHRQ report established their own approach. They assigned points if the study met predefined criteria for quality (a total of 44 points for 18 criteria). They decided to exclude the 20% of articles with the lowest quality scores from further analysis.

Other publications have addressed the question of how to assess research studies on diagnostic testing. The *Journal of the American Medical Association* has published numerous user guides to assist clinician recognition of high-quality clinical research. The guides for understanding diagnostic testing list several important criteria for judging the validity of the results of a study.[64,65] The primary guides are as follows: (1) there was an independent, blind comparison with a reference standard; and (2) the patient sample included an appropriate spectrum of patients to whom the test was applied. The secondary guides include the following: (1) the results of the test being evaluated did not influence the decision to perform the reference standard (verification bias); and (2) the methods for performing the test were described in sufficient detail to permit replication. Although the articles in the *Journal of the American Medical Association* are useful guides for clinicians, they do not provide a methodology or scoring system for rating studies. The approach published by Sackett et al[9] used the following small but essential number of study design features for rating research on diagnostic tests including: (1) an independent blind comparison with a reference standard; (2) an appropriate spectrum of consecutively referred patients (*ie*, avoidance of selection bias); and (3) the use of a reference standard applied to all study patients. These criteria have been organized into levels of evidence (I through V), with level I evidence considered to be the best.[9]

The ATS/AASM/ACCP working group elected not to follow the methodology of the AHRQ review on sleep apnea for rating the quality of research evidence because of several concerns. First, the rationale for assigning some criteria a large number of points (*eg*, randomized controlled trial design, 10 points; study test readers blind to clinical status, 5 points) and others a small number of points (*eg*, verification bias [results of the study test do not determine who gets a polysomnogram], 1 point; patients included with a wide spectrum of sleep apnea severity, 1 point) was not clear. Second, they included a criterion for study quality (*ie*, randomized control design) that is not listed as a criterion in either the series in the *Journal of the American Medical Association* or the article by Sackett et al,[9] and they provided a score that was 10 times greater than other important quality criteria such as avoidance of verification bias.[7] There is an important distinction between a randomized controlled trial and the random assignment of subjects to the order of having either the portable monitoring test or polysomnography first. The latter may be important if testing is likely to have an order effect. For example, if there is a first-night effect for polysomnography that is different from that of portable monitoring testing, this could influence the results of the comparison. The ATS/AASM/ACCP working group adapted the method proposed by Sackett et al[9] to rate the level of evidence of the articles included in this systematic review (see section 2.3).

Subsequent to the ATS/AASM/ACCP working group completing its evaluation of the literature on portable monitoring for sleep apnea, the AHRQ issued a report entitled "Systems to Rate the Strength of Scientific Evidence."[66] The report outlines the following five key domains and elements for systems to rate the quality of individual articles on diagnostic test studies: (1) study population; (2) adequate description of the test; (3) appropriate reference standard; (4) blinded comparison of the test and the reference; and (5) avoidance of verification bias. We assume that the first domain (study population) refers to an appropriate spectrum of

consecutively referred patients (*ie*, avoidance of selection bias). Therefore, the criteria published by Sackett et al[9] include four of these elements (elements 1, 3, 4, and 5). To be part of this current systematic review on portable monitoring for investigating sleep apnea, the results had to be compared to polysomnography. It was not possible to rate the different methods/definitions used for performing polysomnography. Each article was considered equal in this regard. The working group assessed whether there was an adequate description of the methods used to record the signals used by the portable monitor, to define an event (including definitions of events), and whether the method used to score the event was properly described (section 2.3.2). Thus, these two quality criteria are reflected in the quality rating score. Of 30 studies that were rated as having evidence level I or II, only 1 study failed to meet both criteria and had a quality rating of "d" (Table 2). Seven other articles failed to meet one of these criteria, one of which had a quality rating of d, while three each were rated a or b. We do not believe that incorporating the domain of "adequate description of the test" into the evidence level would have affected the results or conclusions of this report.

In addition to defining important domains for systems to rate the quality of evidence for randomized clinical trials, observation studies and diagnostic test studies, the AHRQ report[7,8] also has listed the important domains for systematic reviews and for systems for grading the strength of a body of evidence. For systematic reviews, the AHRQ report recommends that the following 11 domains be addressed:

1. Study question;
2. Search strategy;
3. Inclusion and exclusion criteria;
4. Interventions;
5. Outcomes;

Table 2—*Number of Studies Published on Use of Portable Monitors for Diagnosing Sleep Apnea*

Monitor Type	Study Location	Evidence Level	Quality Rating				References	Total Studies, No.
			a	b	c	d		
2	Sleep laboratory	I						
		II	1				25	
		IV	1			1	19* / 32†	
	Subtotal							3
	Home	I						
		II						
		IV	1	1			19* / 20	
	Subtotal							2
	Total							5
3	Sleep laboratory	I	3				21, 22, 31	
		II	4	1			18‡, 23, 28, 29* / 27	
		IV				1	60	
	Subtotal							9
	Home	I						
		II	2				26, 29*	
		IV	1	1			24 / 30	
	Subtotal							4
	Total							13
4	Sleep laboratory	I	1	1			36 / 45	
		II	10	3	1	1	10, 11, 12, 33,* 34, 43, 44, 52, 54, 57 / 13, 37, 55 / 48 / 35	
		IV	2	2	3	5	15, 38* / 14, 58 / 16, 47, 49 / 46,† 17, 51,* 50 59	
	Subtotal							29
	Home	I		1		1	53 / 40	
		II	1		1		33* / 41	
		IV	1	1	2	1	38* / 39 / 42, 56, 46 / 51*	
	Subtotal							9
	Total							38

*Study included comparisons of in-laboratory polysomnography to portable monitoring performed both in the sleep laboratory and at home.
†Unclear of the exact location of the study; classified as sleep laboratory.
‡Study compared in-laboratory polysomnography to portable monitoring performed both in the sleep laboratory and at home but combined into one group for analysis. Classified as sleep laboratory as the majority of portable studies included in the group were performed in the sleep laboratory.

6. Data extraction;
7. Study quality and validity;
8. Data synthesis and analysis;
9. Results;
10. Discussion;
11. Funding

In this systematic review, we have outlined the study questions (sections 1.0, 4.1, and 4.2) and the search strategy, including inclusion and exclusion criteria (section 2.1). We reviewed portable monitoring for sleep apnea that is performed in a sleep laboratory (*ie*, attended setting, usually performed simultaneously with polysomnography) and/or at home (*ie*, unattended setting). The outcomes reviewed are diagnostic accuracy (*ie*, sensitivity, specificity, and LRs), the results of testing on a population of patients suspected of having sleep apnea (*ie*, the percentage of patients with a result that is either positive or negative and the percentage of those who are misclassified by the portable monitor), as well as cost, failure rate, and repeatability. A description of data extraction (section 2.2), and a detailed process for evaluating study quality and validity (section 2.3) have been included. Data from 51 studies have been synthesized, and the results have been presented in a series of tables. Although a formal meta-analysis was not performed, the summaries provided in these tables make it possible to determine how much evidence there is regarding a particular question, how "good" the evidence is, and whether there was a consistent finding among higher quality studies. Recommendations for improving the quality of research on diagnostic methods for sleep apnea are discussed in section 5.0.

2.0. Materials and Methods

2.1. Literature Review

The ATS/ACCP/AASM working group contracted with the RTI-UNC evidence practice center to conduct a systematic review of the literature and to abstract data in a standard fashion from relevant studies that allowed summaries of their findings to be generated by the ERC. The RTI-UNC team followed the recommended methods for conducting systematic reviews, which emphasized comprehensive literature search and evaluation, and used standardized procedures for the review (and its documentation) of selected articles.

A systematic review of the literature on the diagnosis of sleep apnea was completed in 1997 by the AHRQ.[7] Our literature search focused on articles published since 1997. The initial search was completed June 26, 2001. The bibliographies from two American Sleep Disorder Association reviews[4,6] also were searched for relevant articles. Several search strategies were used, focusing on screening, diagnosis, and costs. The search strategy used the headings "Screening" (including the terms "Reproducibility of Results," or "Predictive Value of Tests," or "Sensitivity and Specificity") "Diagnosis" for finding citations involving the terms "Sleep Apnea Syndromes," "Sleep Apnea (Obstructive)," "Oximetry," "Polysomnography," "Monitoring Physiologic," "Airway Resistance," "Upper Airway Resistance Syndrome," "Respiratory Disturbance Index," "Autoset," "Snoring," or "Respiratory Event-Related Arousals." The term "Home Care Services" also was used to identify citations. For the heading "Screening," the MEDLINE search identified 157 citations, and for "Diagnosis" the MEDLINE search identified 180 citations. The use of the terms "Home Care Services" and "Polysomnography" identified 14 additional citations.

For costs, the MESH heading "Costs and Cost Analysis" was exploded to include the terms "Cost Benefit Analysis," "Cost Allocation," Cost Control," "Cost Savings," "Cost Sharing," "Cost of Illness," "Health Care Costs," and "Health Expenditures." The MEDLINE search was conducted from 1997 to the present and identified 35 citations.

The inclusion criteria were as follows:

- Male/female patients, ages ≥ 18 years, with ANY diagnosis of obstructive sleep apnea;
- Study published in English, no race or gender restrictions;
- Portable device used for diagnosis;
- Polysomnography or other acceptable objective test used for the diagnosis of sleep apnea;
- After completion of the study, each analysis group was ≥ 10 subjects;

The exclusion criteria were as follows:

- Studies in children;
- Studies in languages besides English;
- Reviews, meta-analyses, case reports, abstracts, letters, and editorials.

The titles of retrieved articles were reviewed, and the abstract of any article the title of which mentioned *diagnosis of sleep apnea* was reviewed for relevance to this review. If there was ambiguity about the study meeting the inclusion/exclusion criteria, the full article was reviewed. The reference lists of articles included in this review were scanned to determine other possible articles that should be included.

2.2. Evidence Tables

RTI-UNC worked closely with the ERC to identify the key questions, to develop an abstract review form, to identify the key extraction elements, and then to develop a data extraction elements form (see "Appendix"; available online at http://www.chestjournal.org/cgi/content/full/124/4/1543/DC1). Two evidence practice center reviewers then abstracted complete data independently from each study. The reviewers then compared their results for each element on the data extraction form for each study, and in situations in which there was disagreement a consensus was reached among the reviewers. The final data abstraction forms then were completed by the evidence practice center and were sent to the members of the ERC, who decided on how the evidence would be summarized. The ERC elected to have the search updated to include articles up to December 31, 2001; that identified two additional articles, which members of the ERC abstracted.

2.3. Rating the Quality of Research Articles

The assessment of the study quality was performed by rating 10 separate features of each article that allowed categorization of the evidence level of an article as level I, II, III, or IV (based on three of these items), and then by using a further rating of study quality (a, b, c, or d) that was based on the remaining items. Each of the 10 items was independently rated by two RTI-UNC reviewers

and by two members of the ERC. A final evidence level and quality rating was determined by consensus of the ERC, based on preselected standard definitions.

2.3.1. Evidence level (I, II, III, and IV)

The ERC relied on the presence or absence of three key indicators of quality that dictated the assignment of evidence level based on an approach published by Sackett et al.[9] The definitions of these evidence levels are listed below as follows:

I, blinded comparison, consecutive patients, reference standard performed on all patients;
II, blinded comparison, *nonconsecutive patients*, reference standard performed on all patients;
III, blinded comparison, consecutive patients, reference standard *not performed* on all patients;
IV, reference standard was not applied blindly or independently.

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

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

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

Reference standard was performed on all patients: all patients entered into the study must have undergone both a portable monitor test and a polysomnogram. If the results of one test influenced the decision to perform the other, then this criterion was deemed not to have been met.

2.3.2. Quality rating (a, b, c, d)

Seven other aspects of a study's methodology were scored, and a quality rating was assigned based on the number of indicators for which the study met the criteria. Although the random assignment of testing was an important indicator, it was not applicable to studies that had studied a portable monitor simultaneously with polysomnography. Thus, in some circumstances studies were rated on six indicators rather than seven. The quality indicator (a to d) was based on the number of indicators for which that study did not meet the criteria, as follows:

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

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

1. *Prospective recruitment of patients*: the portable monitoring test and the polysomnogram were performed as patients were recruited into the study rather than reviewing a series of patients who had previously been studied.
2. *Random order of testing*: patients were assigned to undergo portable monitoring testing or polysomnography first at random rather than at the discretion of the investigators. If the portable monitoring study was performed simultaneously with the polysomnogram, this indicator was not rated.
3. *Low data loss (< 10%)*: there were < 10% of patients whose results could not be compared because of the loss of polysomnography or portable monitoring data.
4. *High percentage completed (> 90%)*: of the patients who were initially enrolled into the study (not counting *a priori* exclusions), > 90% completed the study protocol.
5. *Polysomnography methodology/definitions fully described*: the polysomnography methods must include the following:
   a. *characterization of the equipment used*;
   b. *definitions and criteria of all types of breathing events scored and used in comparisons*.
6. *Portable monitor methodology/definitions fully described*: the polysomnography methods must include the following:
   a. *characterization of the equipment used*;
   b. *definitions and criteria of all types of breathing events scored and used in comparisons*
7. *Portable monitor scoring fully described*: includes a clear statement of whether manual or automated scoring was used, and, if automated, whether there was manual review/revision done.

2.4. Approach to Summarizing the Evidence on Portable Monitors

When evaluating the diagnostic accuracy of portable monitors, almost all studies chose to report results as sensitivity and specificity. Many studies examined multiple thresholds for defining a positive result that gave combinations of sensitivity and specificity. When trying to address the issue of whether a portable monitor could reduce the probability that a patient had sleep apnea (section 4.1.1), articles were examined for their best-reported sensitivity, since this should provide the lowest number of false-negative results and the lowest LRs. In circumstances in which various combinations of sensitivity and specificity were reported and two sensitivities were close in value, the "best-reported sensitivity" was taken as the value with the higher corresponding specificity. If the authors had reported different definitions for sleep apnea, the working group selected an AHI definition of ≤ 15, when it was reported, because it was thought that most clinicians would want to know the probability that a patient had an AHI below this level, since theoretically they would then be making a decision on whether or not to offer a trial of therapy. Conversely, when the working group addressed the issue of whether a portable monitor could increase the probability of sleep apnea, the research studies were examined for their best-reported specificity, since this should provide the highest LR and the lowest number of false-positive results. In circumstances in which various combinations of specificity and sensitivity were reported, and two specificities were close in value, the "best-reported specificity" was taken as the value with the higher corresponding sensitivity.

When the best-reported sensitivity and the best-reported specificity used different thresholds (*ie*, different points on a ROC curve), then some patients would meet one or the other criteria, but some would meet neither and therefore would have an indeterminate result. It is important to examine the percentage of patients who meet the criteria for a negative result (*ie*, best-reported sensitivity) and the percentage who meet the criteria for a positive result (*ie*, best-reported specificity), which are reported for each study in Tables 3 and 4, respectively. It is also important to determine the percentage of patients meeting the criteria who had a false result (*ie*, were misclassified by the diagnostic test). This information is reported in Tables 3 and 4, and will be affected by the prevalence as well as the operating characteristics of the test (*ie*, sensitivity, specificity, and LRs). The number of studies, summarized by their level of LRs, monitor type, study location, and evidence level for the best-reported sensitivity and the best-reported specificity, are pre-

Table 3—*Best-Reported Sensitivity and Calculated Negative LRs

Monitor Type	Study Location	Evid Level	Quality Rating	First Author	Reference	Pts, No.	Prev	Neg LR and CIs			Best Sens and CIs/Corr Spec				AHI Defn	Pts with Neg PM Result, %	False Neg, %
								LR	Low CI	Up CI	Sens	Low CI	Up CI	Spec			
2	Home	IV	b	Portier F	20	78	47.0	0.19	0.10	0.38	81	68.3	93.7	98	15	61	15
	Sleep laboratory	II	a	Mykytyn I	25	20	50.0	0.22	0.06	0.78	80	55.2	100	90	10	55	18
		IV	d	Orr WC	32	40	62.5	†	‡	‡	100	‡	‡	93	15	35	0
3	Home	II	a	White D	29	70	61.4	0.13	0.05	0.34	91	82.4	99.6	70	10	32	17
				Parra O	26	89	84.3	0.15	0.04	0.52	95	90.1	99.9	33	10	9	45
		IV	a	Ancoli-Israel S	24	34	73.5	†	‡	‡	100	‡	‡	66	10	18	0
			b	Whittle A	30	58	55.2	0.43	0.22	0.85	75	60.0	90.0	58	15	40	35
	Sleep laboratory	I	a	Zucconi M	21	29	65.5	†	‡	‡	100	‡	‡	100	10	34	0
				Emsellem H	22	63	61.9	0.05	0.01	0.21	95	88.2	100	96	5	40	8
				Verse T	31	53	47.2	0.08	0.02	0.3	92	81.8	100	96	10	54	7
		II	a	White D	29	30	63.3	†	‡	‡	100	‡	‡	64	10	24	0
				Ficker J	28	51	56.9	0.03	0.00	0.24	97	90.8	100	100	10	45	4
				Redline S	18	25	84.0	0.05	0.01	0.32	95	85.7	100	100	10	20	21
				Man G	23	104	26.9	0.15	0.06	0.37	86	73.1	98.9	95	15	73	5
			b	Ballester E	27	116	24.1	0.12	0.04	0.34	89	77.4	100	92	10	72	4
		IV	d	Claman D	60	42	50.0	0.15	0.05	0.43	86	71.2	100	95	15	54	13
4	Home	I	b	Series F	53	240	45.8	0.04	0.01	0.16	98	95.4	100	48	10	27	3
			d	Gyulay S	40	98	43.9	0.14	0.04	0.42	93	85.4	100	51	15	32	10
		II	a	Baltzan M	33	66	41.2	0.31	0.09	1.06	90	78.7	100	32	15	23	18
			c	Golpe R	41	116	61.2	0.33	0.18	0.62	84	75.5	92.5	48	10	28	34
		IV	b	Schafer H	39	114	70.2	0.12	0.04	0.34	95	90.2	99.8	41	10	16	22
			c	Williams A	42	36	55.5	0.35	0.19	0.64	65	44.1	85.9	100	10‖	64	30
				Ryan P	56	69	46.3	0.69	0.55	0.87	31	15.0	47.0	100	15	86	37
			d	Olson L	51	793	44.0	0.17	0.09	0.31	97	95.2	98.8	18	15	11	12
	Sleep laboratory	I	a	Vazquez J	36	241	59.0	0.04	0.01	0.10	97	94.2	99.8	80	10	35	5
			b	Chiner E	45	275	77.7	0.24	0.17	0.33	82	76.8	87.2	76	15	31	45
		II	a	Koziej M	44	56	66.1	†	‡	‡	100	‡	‡	63	10	21	0
				Esnaola S	43	150	60.0	0.03	0.01	0.11	98	95.1	100	78	10	32	4
				Gugger M	11	67	46.3	0.04	0.01	0.29	97	91.0	100	77	20	43	3
				Mayer P	12	95	93.7	0.06	0.01	0.25	97	93.5	100	50	5	6	47
				Stoohs R	34	56	46.4	0.08	0.02	0.30	92	81.6	100	97	10	56	7
				Baltzan M	33	86	41.2	0.09	0.01	0.64	97	91.4	100	32	15	20	6
				Zamarron C	52	233	53.2	0.09	0.05	0.19	94	89.8	98.2	65	10	34	10
				Issa F	57	129	54.3	0.12	0.06	0.23	89	81.7	96.3	95	7	49	12
				Douglas N	54	200	45.5	0.36	0.27	0.48	67	57.3	76.7	92	15	65	23
			b	Kiely J	13	36	36.1	†	‡	‡	100	‡	‡	92	15	59	0
				Yamashiro Y	37	269	50.9	0.08	0.04	0.16	94	90.0	98.0	74	5	39	8
				Rauscher H	55	116	40.5	0.10	0.03	0.30	94	87.2	100	62	10	39	6
			c	Cooper B§	48	41	29.2	0.29	0.11	0.78	75	50.5	99.5	86	15	68	11
			d	Gurubhagavatula I	35	359	69.4	0.15	0.11	0.21	86	81.7	90.3	91	5	38	26
		IV	a	Bagnato M	15	56	78.6	0.05	0.01	0.44	98	93.9	100	40	10	10	16
			b	Bradley P	14	31	48.4	†	‡	‡	100	‡	‡	92	15	48	0
				Zamarron C	58	197	46.2	0.21	0.14	0.32	81	72.9	89.1	91	10	58	15
			c	Levy P	49	301	64.1	0.04	0.02	0.12	98	96.0	100	46	15	18	7
				Gugger M	16	27	70.3	0.20	0.07	0.54	82	64.7	99.3	90	20	39	32
			d	Fleury B	17	38	46.9	†	‡	‡	100	‡	‡	87	10‖	46	0
				Pepin J	50	26	57.7	0.11	0.02	1.11	95	84.0	100	45	5	22	13
				Olson L	51	113	22	0.11	0.03	0.42	92	81.3	100	73	15	59	3
				Bonsignore G	46	83	56.6	0.26	0.16	0.42	75	62.0	87.0	100	10	58	25

*Total of 49 studies from 46 research studies. Five studies were not included (15, 33, 35, 37, 47) because sensitivity/specificity data were not reported. Pts = patients; Prev = prevalence; Neg LR = negative LR; CI = confidence intervals; Low CI = lower confidence interval; Up CI = upper confidence interval; Best Sens = the best sensitivity reported in the results of the article (for an AHI cutoff of ≤ 15, unless not reported) that produced the lowest calculated LR; Corr Spec = the specificity of the test result that corresponded to the best-reported sensitivity; Sens = sensitivity; Spec = specificity; AHI Defn = the AHI cutoff (from polysomnography) that was used to define patients with sleep apnea at the best-reported sensitivity. See Table 1 for other abbreviations not used in the text.

†Total is 0 because sensitivity = 100%.

‡Cannot be calculated.

§Not performed simultaneously with PSG.

‖Apnea index, not AHI.

Table 4—*Best-Reported Specificity and Calculated Positive LRs

Monitor Type	Study Location	Evid Level	Quality Rating	First Author	Reference	Pts, No.	Prev	Pos LR and CIs			Best Spec and CIs/Corr Sens				AHI Defn	Pts With Pos PM Result, %	False Pos, %
								LR	Low CI	Up CI	Spec	Low CI	Up CI	Sens			
2	Home	IV	b	Portier F	20	78	47	40.5	4.8	343	98	93.7	100	81	15	39	3
	Sleep laboratory	II	a	Mykytyn I	25	20	50	8	1.2	52.4	90	71.4	100	80	10	45	11
		IV	d	Orr WC	32	40	62.5	14.3	2.3	90.4	93	80.1	100	100	15	65	4
3	Home	II	a	Parra O	26	89	84.3	9	1.3	61.0	93	79.6	100	63	10	54	2
				White D	29	70	41.4	5.1	2.2	11.7	83	68.8	97.2	86	20	46	22
		IV	a	Ancoli-Israel S	24	34	73.5	2.9	1.2	7.3	66	35.1	96.9	100	10	82	11
			b	Whittle A	30	58	55.2	1.8	1.1	2.9	58	39.0	77.0	75	15	60	31
	Sleep laboratory	I	a	Zucconi M	21	29	65.5	†	‡	‡	100	‡	‡	100	10	66	0
				Verse T	31	53	47.2	23.3	3.7	141	96	88.7	100	92	10	46	5
				Emsellem H	22	63	61.9	23.8	3.3	168	96	88.2	100	95	5	60	2
		II	a	Ficker J	28	51	56.9	†	‡	‡	100	‡	‡	97	10	55	0
				Redline S	18	25	84	†	‡	‡	100	‡	‡	95	10	80	0
				Man G	23	104	26.9	17.2	6.4	46.3	95	90.1	99.9	86	15	27	14
				White D	29	30	43.3	6.4	1.3	32.3	88	68.8	100	77	20	40	17
			b	Ballester E	27	116	24.1	11.1	5.4	22.8	92	86.3	97.7	89	10	28	22
		IV	d	Claman D	60	42	50	17.2	2.6	112	95	85.7	100	86	15	46	6
4	Home	I	b	Series F	53	240	45.8	1.9	1.6	2.2	48	39.4	56.6	98	10	73	39
			d	Gyulay S	40	98	43.9	20	3.1	129	98	94.3	100	40	15	18	6
		II	a	Baltzan M	33	66	41.2	10.3	1.7	63.5	97	91.6	100	31	15	14	12
			c	Golpe R	41	116	61.2	10.7	2.0	56.8	97	92.0	100	32	10	21	6
		IV	b	Schafer H	39	114	70.2	4.0	1.8	9.1	85	73.0	97.0	60	10	47	10
			c	Williams A	42	36	55.5	†	‡	‡	100	‡	‡	65	10§	36	0
				Ryan P	56	69	46.4	†	‡	‡	100	‡	‡	31	15	14	0
			d	Olson L	51	793	44	1.8	1.6	2.0	54	49.4	58.6	83	5	62	41
	Sleep laboratory	I	a	Vazquez J	36	241	49	8.2	5.1	13.2	88	82.3	93.7	98	15	54	11
			b	Chiner E	45	275	77.7	8.9	3.5	22.1	93	86.6	99.4	62	15	50	3
		II	a	Issa F	44	129	31	45	10.5	193	98	95.1	100	90	20	29	5
				Esnaola S	43	150	60	34.5	5.8	203	98	94.5	100	69	10	42	2
				Stoohs R	34	56	46.4	30.7	4.0	235	97	90.9	100	92	10	44	4
				Douglas N	54	200	45.5	15.8	6.0	51.6	97	93.8	100	52.7	10	26	6
				Baltzan M	33	86	41.2	14	2.0	98.4	98	94.1	100	28	15	13	9
				Mayer P	12	95	54.7	11.2	3.8	33.8	93	85.4	100	79	30	46	7
				Gugger M	11	67	46.3	4.2	2.3	7.7	77	63.3	90.7	97	20	57	22
				Zamarron C	52	233	53.2	3.6	2.3	5.6	84	77.1	90.9	57	10	38	20
				Koziej M	44	56	66.1	2.7	1.5	4.9	63	41.3	84.7	100	10	79	16
			b	Kiely J	13	36	36.1	12.5	3.1	50.0	92	80.9	100	100	15	41	12
				Yamashiro Y	37	269	50.9	3.6	2.7	4.8	74	66.5	81.5	94	5	61	21
				Rauscher H	55	116	40.5	2.5	1.8	3.4	62	50.5	73.5	94	10	61	37
			c	Cooper B‖	48	41	48.8	12	1.8	79.0	95	85.7	100	60	5	32	8
			d	Gurubhagavatula I	35	359	33	18	10.3	31.3	95	92.2	97.8	90	30	33	10
		IV	a	Bagnato M	15	56	50.7	2.4	1.5	3.8	60	41.7	78.3	97	20	69	29
			b	Bradley P	14	31	48.4	12.5	2.4	65.9	92	78.7	100	100	15	52	8
				Zamarron C	58	197	46.2	9.0	4.9	16.6	91	85.6	96.4	81	10	42	12
			c	Levy P	49	301	64.1	†	‡	‡	100	‡	‡	46	15	30	0
				Gugger M	16	27	70.3	8.2	1.0	66.0	90	69.2	100	82	20	61	5
			d	Pepin J	50	26	57.7	†	‡	‡	100	‡	‡	50	5	29	0
				Bonsignore G	46	83	56.6	†	‡	‡	100	‡	‡	75	10	42	0
				Fleury B	17	38	46.9	7.7	2.2	26.3	87	71.0	100	100	10§	54	13
				Olson L	51	113	22	6	1.3	27.1	98	95.1	100	12	15	4	37

*Total of 49 studies from 46 research studies. Five studies were not included (15, 33, 35, 37, 47) because sensitivity/specificity data were not reported. Pos LR = LR if the portable monitor test was positive for sleep apnea; Best Spec = the best specificity reported in the results of the paper that produced the highest calculated LR; Corr Sens = the sensitivity of the test result that corresponded to the best-reported specificity. See Tables 1 and 3 for abbreviations not used in the text.

†Total is infinity because specificity = 100%.

‡Cannot be calculated.

§Apnea index, not AHI.

‖Not performed simultaneously with PSG.

sented in Tables 6 and 7, respectively (available online at http://www.chestjournal.org/cgi/content/full/124/4/1543/DC1). A similar summary for articles with both high and low LRs is presented in Table 8 (available online at http://www.chestjournal.org/cgi/content/full/124/4/1543/DC1).

Ideally, a portable monitor test would have a single cutoff that has both a high sensitivity and a high specificity, so that patients are either negative or positive and there is no "gray zone." To address the question of whether some monitors came close to being an "ideal test" that had thresholds that minimized the number of patients in the gray zone and the misclassification rate, studies were examined that had both a high and low calculated LR (results are reported in Table 5 and are summarized in section 4.1.3). As with Tables 3 and 4, Table 5 highlights the percentage of patients with a false result and the percentage of patients who did not meet the criteria for a positive or negative result (*ie*, those in the gray zone). If a study had a single threshold for best-reported sensitivity and specificity, then the percentage of patients without a negative or positive result is 0.

The 95% confidence intervals are reported in Tables 3 and 4 for the best-reported sensitivity and best-reported specificity, respectively, as well as for the corresponding LRs. The 95% confidence intervals were calculated from the reported sensitivity, specificity, prevalence, and number of patients according to the method of Simel et al.[67] In some studies, the prevalence was not reported but was estimated from figures.

When assessing the evidence on portable monitoring, it is important to keep in mind the following points to avoid misinterpreting the data:

- How low does the best-reported sensitivity (*ie*, low LR) reduce the probability of sleep apnea?
- How high does the best-reported specificity (*ie*, high LR) increase the probability of sleep apnea?
- Can the portable monitor both substantially reduce the probability that a patient has sleep apnea (if the test result is negative) and increase the probability that a patient has sleep apnea (if the test result is positive)?
- What percentage of patients in the study actually met the thresholds for a positive or negative test result?
- What percentage of patients who met the thresholds were misclassified (*ie*, had a false result)?
- How likely is it that the results of a study are valid (*ie*, evidence level and quality rating)?
- How precise are the estimates (*ie*, width of the confidence intervals) for sensitivity, specificity, and LRs?
- Were the RDI thresholds used to define a positive and negative result preselected or determined by *post hoc* (retrospective) analysis?

Using different RDI thresholds for a positive or negative result, and using different thresholds to define best-reported sensitivity and specificity, can be difficult to understand. A detailed discus-

Table 5—*Studies With Both High (> 5.0) and Low (< 0.2) LRs**

							Negative PM Result				Positive PM Result					
Monitor Type	Study Location	Evid Level	Quality Rating	First Author	Reference	Pts, No.	Neg LR	Best Sens	Corr Spec	Pts With Neg PM Result, %	Pos LR	Best Spec	Corr Sens	Pts With Pos PM Result, %	Pts With False Pos or False Neg Result, %	Pts Without a Neg or Pos Result, %
2	Home	IV	b	Portier F	20	78	0.19	81	98	61	40.5	98	81	39	10	0
	Sleep laboratory	IV	d	Orr WC	32	40	†	100	93	35	14.3	93	100	65	3	0
3	Home	II	a	White D	29	70	0.13	91	70	32	5.1	83	86	46	16	22
				Parra O	26	89	0.15	95	33	9	9	93	63	54	5	37
	Sleep laboratory	I	a	Zucconi M	21	29	†	100	100	34	‡	100	100	66	0	0
				Emsellem H	22	63	0.05	95	96	40	23.8	96	95	60	5	0
				Verse T	31	53	0.08	92	96	54	23.3	96	92	46	6	0
		II	a	White D	29	30	†	100	64	24	6.4	88	77	40	7	36
				Ficker J	28	51	0.03	97	100	45	‡	100	97	55	2	0
				Redline S	18	25	0.05	95	100	20	‡	100	95	80	4	0
				Man G	23	104	0.15	86	95	73	17.2	95	86	27	7	0
			b	Ballester E	27	116	0.12	89	92	72	11.1	92	89	28	9	0
		IV	d	Claman D	60	42	0.15	86	95	54	17.2	95	86	46	10	0
4	Home	I	d	Gyulay S	40	98	0.14	93	51	32	20	98	40	19	4	49
	Sleep laboratory	I	a	Vazquez J	36	241	0.04	97	80	35	8.2	88	98	54	8	11
		II		Esnaola S	43	150	0.03	98	78	32	34.5	98	69	42	2	26
				Mayer P	12	95	0.06	97	50	6	11.2	93	79	46	6	48
				Stoohs R	34	56	0.08	92	97	56	30.3	97	92	44	5	0
				Baltzan M	33	86	0.09	97	32	20	14	98	28	13	2	67
				Issa F	57	129	0.12	89	95	49	45	98	90	29	7	22
			b	Kiely J	13	36	†	100	92	59	12.5	92	100	41	5	0
			d	Gurubhagavatula I	35	359	0.15	86	91	38	18	95	90	33	13	29
		IV	b	Bradley P	14	31	†	100	92	48	12.5	92	100	52	4	0
			c	Levy P	49	301	0.04	98	46	18	‡	100	46	30	1	52
				Gugger M	16	27	0.20	82	90	39	8.2	90	82	61	16	0
			d	Fleury B	17	38	†	100	87	46	7.7	87	100	54	7	0
				Pepin J	50	26	0.11	95	45	22	‡	100	50	29	3	49
				Olson L	51	113	0.11	92	73	59	6	98	12	4	3	37

*Total of 28 studies from 27 research studies. See Tables 1 and 3 for abbreviations not used in the text.
†Total is 0 because sensitivity = 100%.
‡Total is infinity because specificity = 100%.

sion and example of the effect of varying RDI thresholds, and thus the best-reported sensitivity/specificity, on the number of nondiagnostic and false-positive/false-negative results can be found in an accompanying paper (see page 1535).

3.0. Literature Search Results

3.1. Number of Articles Reviewed/Rejected

The initial literature search resulted in 59 original research articles being identified that met the inclusion criteria. Of these, 46 articles were selected for review by the evidence practice center.[10–27,29–34,36–57] Thirteen articles were not included[68–80] for the following reasons: they were reports of older monitors for which more recent research had been published (one article) or were known to no longer be commercially available (one article); they evaluated technology that was not portable (*eg*, the static charge-sensitive bed) [five articles]; they studied technology that was not widely used or available (four articles); they did not use technology that was involved with monitoring a physiologic signal (one article); or patients had been tested following a surgical intervention (one article).

The search was first extended to October 31, 2001, which identified an additional three articles,[28,58,60] and then to December 31, 2001, which yielded in additional two articles[35,59] for a total of 51 articles included in this review.

3.2. Type of Monitor/Home- vs Sleep Laboratory-Based Studies/Evidence Rating

There were four studies describing five groups of patients published on type 2 monitors (comprehensive polysomnography) [Table 2]. Three of the four studies were rated as having level IV evidence, and one study was rated as having level II evidence.

There were 12 published studies on type 3 monitors, describing 14 groups of patients. There were five assessments of type 3 monitors at home and nine assessments in the sleep laboratory. Overall, the studies were of higher quality (at home, three level II studies; in the sleep laboratory, three level I studies and five level II studies) than the studies of type 2 monitors (Table 2).

The majority of the published studies were on type 4 monitors (35 of the 51 studies), with 29 reports of patients studied in the sleep laboratory and 9 reports of patients studied at home. Approximately 50% of the studies on type 4 monitors had level I or II evidence (Table 2).

3.3. Sensors to Detect Breathing

The categorization of portable monitors according to type may not be completely relevant. It is clear that type 2 monitors are different from other monitors because of their ability to measure EEG and EMG signals. The distinction between type 3 and type 4 monitors is less clear. Regardless of monitor type, usually only one signal is used to define a breathing event; occasionally a second signal is used, and rarely a third channel is used. However, all type 3 monitors have the option of using more channels, and many incorporate a body-position sensor. Many type 4 monitors use only one channel. It may be more important to distinguish between the types of signals that are monitored (*eg*, flow measured by thermistor, flow measured by nasal pressure, or oxygen saturation) and how they are used to detect breathing disturbances. Table 1 groups studies based on the primary signal that was used to define breathing disturbances during sleep. It also indicates how events were defined by polysomnography, differences in oxygen saturation sampling frequency (when oxygen saturation was used to define an event), as well as the type of portable monitor scoring that was used (*ie*, automated, manual, or a combination).

The most common signal used in portable monitors was airflow measured by a thermistor. Ten of 15 studies that used this as the primary scoring channel were rated as having level I or II evidence. Only two of these used the same criteria for defining a breathing event (Table 1). Flow measured by nasal pressure was used in eight studies (four were rated as having level II evidence). All four of the level I/II studies used the same portable monitor and the same criterion for defining a breathing event. An oxygen saturation signal was used in 22 studies as the primary signal to define a breathing event, 13 of which were rated as having level I or II evidence. In these 13 studies, there were 11 different criteria used for scoring an event. More studies using flow, measured by nasal pressure or oxygen saturation, utilized automatic scoring than did studies that used flow measured by thermistor, which used manual scoring or a combination.

3.4. Bland and Altman Analysis

Some studies reported how well the data from portable monitoring agreed with those of polysomnography using mean differences between the two methods and the limits of agreement (Table 9; available online at http://www.chestjournal.org/cgi/content/full/124/4/1543/DC1). A total of 24 studies reported this analysis (type 2 monitors, 2 studies; type 3 monitors, 6 studies; type 4 monitors, 16 studies). A few authors reported confidence intervals rather than limits of agreement (Table 9). The limits of agreement tended to be quite wide, suggesting that the two methods did not agree particularly well.

But the limits were wider for higher levels of RDI, for which it was less important to get the same number as it was at lower levels of RDI. The limits of agreement can be adjusted by using a logarithmic transformation of the differences, but investigators rarely did this. Thus, it is challenging to interpret these data and make a recommendation about the utility of portable monitoring based on this alone.

4.0. Discussion of the Evidence on Portable Monitors

4.1. Primary Questions

There are many causes of variability in results between portable monitors and polysomnography. A simultaneous comparison of portable monitoring with polysomnography (sleep laboratory-attended) controls for a number of conditions that nonsimultaneous studies do not and, as such, provides an estimate of what might ideally be expected during home-unattended portable monitoring. The comparison of portable monitoring performed in the home-unattended setting with sleep laboratory polysomnography may not capture differences that favor one environment (*ie*, home or laboratory) over the other.

Conclusions regarding the utility of portable monitors are most applicable to the population of patients, and the methods that the portable monitors used for detecting events, that are the focus of this report. As detailed in the following sections, the prevalence of sleep apnea was high, averaging about 55%. Patients were predominantly male and generally were selected for studies by practitioners with expertise in evaluating patients with sleep apnea. The methods for scoring respiratory events on polysomnography varied from study to study.

A number of studies examined multiple RDI thresholds for determining an abnormal AHI. These optimal thresholds were determined using *post hoc* analyses and, therefore, may not necessarily be reproducible. LRs calculated from the reported sensitivity and specificity data from many studies had wide 95% confidence intervals, indicating a lack of precision in these estimates. A meta-analysis using summary ROC curves would have allowed for pooled results and narrower 95% confidence intervals. However, the ERC thought that there was not enough similarity between studies to warrant using this approach.

The effect of using time in bed for portable monitors, in general, led to a slightly higher AHI by polysomnography than RDI for type 3 monitors, with inconsistent effects on type 4 monitors (Table 10; available online at http://www.chestjournal.org/cgi/content/full/124/4/1543/DC1). The effect of the higher AHI is, in general, to reduce sensitivity and increase specificity.

4.1.1. Evidence that portable monitors can be used to reduce the probability that a patient has an abnormal AHI

If a portable monitor is going to be used to exclude a diagnosis of sleep apnea, the monitor needs to have a high sensitivity/low LR for a negative result. The percentage of patients who will have a negative result is determined by (1) the pretest probability or prevalence of the disease (a characteristic of the population being studied) and (2) the sensitivity/specificity or negative LR (the operating characteristics) of the portable monitor being used. The research conducted to date has been in sleep clinic populations that have a very high prevalence of disease. The result is that, even by using a portable monitor with very good operating characteristics, a small percentage of patients will be classified as negative with a higher chance for those with a negative result to be classified incorrectly (*ie*, a false-negative result). Details about the studies covered in this section can be found in Tables 3 (page 1557) and 6 (http://www.chestjournal.org/cgi/content/full/124/4/1543/DC1).

4.1.1.1. Type 2 monitors

There were three studies that reported results on sensitivity and specificity.[20,25,32]

4.1.1.1.1. Sleep laboratory-attended

There were two studies performed in the sleep laboratory, with one rated as having level II evidence but reporting on a small number of patients.[25] The calculated LR for a negative result was not very low (0.22), and the confidence intervals were very wide.

4.1.1.1.2. Home-unattended

One study was performed at home[20] but was given a low evidence rating. The study did not report a very low LR for a negative result (0.19), and, of the patients who had a negative result, a substantial proportion of the results (15%) was false-negative.

4.1.1.2. Type 3 monitors

4.1.1.2.1. Sleep laboratory-attended

There were nine studies performed on patients in a sleep laboratory setting where the portable monitoring occurred simultaneously with the polysomnogram. Eight of these studies[18,21–23,27–29,31] were rated as having either level I or II evidence, all of which had very low or reasonably low LRs (range, 0 to 0.15). Two studies[21,29] reported a sensitivity of 100% resulting in a LR of 0, but both had small numbers of patients. All eight studies reported a substantial proportion of patients (range, 20 to 73%) having a negative result on portable monitoring, with a small percentage of those results being false-negative (range, 4 to 8%). In seven of those studies, the

portable monitor used flow measured by thermistor to detect events. Four of the studies included oxygen desaturation as a necessary criterion for hypopneas.

4.1.1.2.2. Home-unattended

There were four studies reporting data on portable monitoring performed in the home setting; two of which[26,29] were rated as having level II evidence. Both of these studies and one of the level IV evidence studies[24] reported low LRs in a modest number of patients. Two of these studies[24,26] had a very high prevalence, so that the number of patients who had a negative result was small (9% and 18%, respectively). The study with a higher percentage of negative results (32%) had a substantial percentage of false-negative results (17% of those with a negative result).[29] All three of these studies used flow measured by thermistor, with two of the studies including oxygen desaturation as a necessary criterion for hypopneas.

The majority of the evidence from attended monitoring indicates that type 3 monitors using flow measured by thermistor can substantially reduce the probability of sleep apnea in a substantial percentage of patients. This approach is not as well-validated in the home setting with this type of monitor.

4.1.1.3. Type 4 monitors

The methods used to analyze type 4 portable monitors were diverse. Devices measured one, two, or three variables and reported on these either individually or in combination, depending on the device (Table 1). The single variables measured included oximetry, heart rate, airflow, and, in one case, esophageal pressure. Methods of analysis used oxygenation criteria ranging from 2 to 5% desaturation, derived oxygen desaturation indexes (*eg*, the "delta index"), the time spent below a certain saturation level (usually 90%), and qualitative visual inspection of the output of the oximeter.

Computer scoring of the portable monitor without editing was used in 16 studies. Manual scoring was used in seven studies (Table 1). In two studies, a combination of computer scoring without editing for one variable and manual scoring for another variable were used. One study demonstrated a marked and significant loss of sensitivity of an oximeter using a 12-s sampling rate compared to a 2-s sampling rate.[38] Three studies reported using a 12-s sampling rate, 10 studies did not report the sampling rate, 4 studies used sampling rates between 2 and 12 s, and 3 studies used a sampling rate of $\leq$ 2 s (Table 1). The sampling rates in the rest of the studies were not applicable due to the signals, such as airflow, that were used to determine the RDI or its equivalent for the portable monitor. The result of using a low sampling rate is potentially to reduce the sensitivity and increase the specificity of the test. All of the studies used time spent in bed as a reference for portable monitor indexes compared to time asleep for polysomnography, which, as previously discussed, would tend to reduce the portable monitor index compared to polysomnography. In two studies[33,44] in which computer scoring was compared to manual scoring, manual scoring was superior.

4.1.1.3.1. Sleep laboratory-attended

There were 16 studies of type 4 monitors that were rated as having either level I evidence[36,45] or level II evidence.[11–13,33–35,37,43,44,48,52,54,55,57] An additional nine studies rated as having level IV evidence were reviewed.[14–17,46,49–51,58] Nine of the 16 studies[34–37,45,48,52,54,55] used oxygen saturation as the primary method for detecting breathing disturbances (Tables 1 and 3), 3 studies used flow measured by nasal pressure,[11–13] 2 studies used oxygen saturation, snoring, and heart rate,[43,44] 1 study used thermistor measured flow,[33] and 1 study used snoring and oxygen saturation.[57] A LR of $\leq$ 0.15 was reported in 13 of the 16 studies that were rated as having level I or II evidence (Table 3). One study rated as having level I evidence had a lower sensitivity (and higher LR for a negative result) of 82% but used a stricter definition of oxygen desaturation than other studies to define an event (*ie*, $>$ 4%).[45] Most studies had a substantial percentage of patients with a negative result and a very low percentage of false-negative results (Table 3). There were two exceptions, with one study[12] having a very high prevalence (93.7%) and the other study[45] having a lower sensitivity. There were no obvious differences in diagnostic performance between monitors using different technologies for detecting events. Studies rated as having level IV evidence also reported reasonably low LRs with no obvious differences between the different techniques used for event detection.

4.1.1.3.2. Home-unattended

There were eight studies reporting sensitivity/specificity data on type 4 portable monitoring performed in the home setting, four of which[33,40,41,53] were rated as having level I or II evidence (Table 3). All eight studies utilized oxygen saturation monitoring (seven studies used it as the primary criterion, and one study used it as a secondary criterion with flow measured by thermistor). Both level I studies reported higher sensitivities than the studies rated as having level II evidence. Two of the level IV studies reported modestly low LRs. The prevalence in most of these studies was between 40% and 70%. The percentage of patients testing negative on the porta-

ble monitor who had falsely negative results was lower in the two level I studies (3% and 10%) than in the two level II studies (18% and 34%) or the level IV studies (12 to 37%).

As with type 3 monitors, the majority of the evidence from attended monitoring indicates that type 4 monitors can substantially reduce the probability of sleep apnea in an important percentage of patients. There are fewer studies in the home setting with more varied results, however the two studies with the highest evidence rating indicate that it is possible to get comparable results in an unattended setting. There does not appear to be any obvious differences in reported results between type 4 monitors that use oxygen saturation compared with flow measured by nasal pressure as signals to detect events. However, nasal pressure monitors have not been tested unattended in the home.

4.1.2. Evidence that portable monitors can be used to increase the probability that a patient has an abnormal AHI

If a portable monitor is going to be used to rule in a diagnosis of sleep apnea, the monitor needs to have a high specificity/high LR for a positive result. Similar to the goal of reducing the probability of sleep apnea (section 4.1.1), the percentage of patients who will have a positive result is determined by (1) the pretest probability or prevalence of the disease (a characteristic of the population being studied) and (2) the specificity, sensitivity, and positive LR (operating characteristics) of the portable monitor. With the high prevalence of disease that is seen in most sleep clinics, more patients in these studies are likely to be classified as being positive by the monitor unless the threshold for a positive result is set very high. Also, because of the high prevalence, patients with a positive result are more likely to have a true-positive result than a false-positive result. Details about the studies covered in this section can be found in Tables 4 and 7.

4.1.2.1. Type 2 monitors

4.1.2.1.1. Sleep laboratory-attended

There were three studies (evidence level IIa,[25] evidence level IV/quality rating b,[20] and evidence level IV/quality rating d[32]) that compared portable monitors simultaneously with polysomnography (Table 4). In the two studies[25,32] that reported sensitivity and specificity (total, 60 patients), specificity was high (90% and 93%, respectively) as were the LRs for a positive result (8 and 14.3, respectively). In one of these studies, 10 patients were studied in the sleep laboratory, but the monitors were unattended by a technician, which led to the loss of airflow and saturation signals for some period of time in some of the patients.[25] Approximately 50% of patients had a positive result, with a range of false-positive results of 4 to 11%.

4.1.2.1.2. Home-unattended

One study was performed at home[20] but was given a low evidence rating. The study reported a very high LR for a positive result (40.5), and, of the patients who had a positive result (39%), a very small proportion (3%) had false-positive results.

4.1.2.2. Type 3 monitors

4.1.2.2.1. Sleep laboratory-attended

There were nine studies (level I or II evidence, eight studies[18,21–23,27–29,31]; level IV evidence, one study[60]) that compared portable monitoring in an attended (sleep laboratory) setting simultaneously with polysomnography recording (Table 4). The studies demonstrated a high degree of specificity, with most values at $> 90\%$. Sensitivity was also high in these studies, with the result that the LRs for a positive result were very high for all of the studies except one.[29] The threshold chosen for the definition of an abnormal AHI had, in general, little effect on the specificity, with one exception in which there was a substantial increase as the AHI threshold was increased.[29] The high positive LRs indicated that a study with a type 3 attended portable monitor is likely to increase substantially the probability of a positive polysomnography study result. These studies used similar scoring definitions in most cases. The exceptions were that arousals were used to define the occurrence of an event on polysomnography in three studies, a compressed time frame was noted to be used in two studies for the portable monitor, computer scoring was used without editing in one study using a portable monitor, and time in bed was used as the reference for the portable monitors in all but one other study. One study[21] used time in bed for both polysomnography and portable monitoring. One study[29] used an indirect measure for sleep with the portable monitor. In three studies[21,28,31] in which manual and computer scoring were compared, manual scoring was reported to be superior. The use of time in bed for the portable monitor may lead to reduced sensitivity and artificially increase specificity.

4.1.2.2.2. Home-unattended

There were four studies published[24,26,29,30] (level II evidence, two studies; level IV evidence, two studies). Specificities ranged from 58 to 93%, and, correspondingly, LRs for a positive result were modest in the two higher quality studies (LRs, 5.1 and 9)[26,29] compared with those in the two lower quality studies (LRs, 1.8 and 2.9).[24,30] The percentage of patients that had a positive result ranged from 46 to

82%, with a corresponding range of false-positive results in those testing positive of 2 to 31%.

Authors commented on several possible biases. Two of the studies in which sleep position was measured at home mentioned that several patients exhibiting false-positive results had experienced more sleep in the supine position at home than with polysomnography in the laboratory. If these false-positive results were, in fact, true-positive results, it can be calculated from the authors' data that the specificity would have increased from 66 to 100% in one study,[24] and from 70.4 to 79.2% at an AHI of < 10 and from 82.9 to 87.2% at an AHI of < 20 in a second study.[29] The data loss was not substantial, except for one study[30] in which between 13% and 18% of studies could not be interpreted.

Studies of type 3 monitors that were performed in an attended setting reported high specificities and produced LRs that could substantially increase the probability that a patient would have an abnormal AHI. A limited number of home, unattended studies showed a wider range of specificities (and sensitivities) with LRs that generally did not markedly improve the probability of sleep apnea with either a positive or negative study result. However, it is possible that some patients labeled as having false-positive results were misclassified because of uncontrollable differences in factors between the at-home and the in-laboratory study, like differences in time spent lying supine. The result is that the reported specificities for unattended studies may be underestimated and the influence of the studies to increase probability is greater than that reported. All studies used flow measured by thermistor to detect events. Two of the four studies[26,29] included an oxygen desaturation criterion as being necessary to measure hypopneas, while two other studies[24,30] did not (Table 1).

4.1.2.3. Type 4 monitors

4.1.2.3.1. Sleep laboratory-attended

There were 25 studies (level I evidence, 2 studies; level II evidence, 14 studies; level IV evidence, 9 studies) performed on type 4 monitors (Tables 4 and 7). There were 13 studies that used oximetry alone,[34–37,45,46,48–52,54,55] 8 studies that used airflow (nasal pressure alone, 6 studies[11–14,16,17]; nasal pressure and oxygen desaturation, 1 study[15]; thermistor with oxygen desaturation, 1 study[33]), 1 study that used snoring with oximetry,[57] 2 studies that used oximetry plus heart rate and snoring,[43,44] and 1 study that used heart rate alone.[58] One other study did not report sensitivity in an interpretable fashion,[59] and three other studies did not report sensitivities and specificities.[10,38,47] Of the 16 studies rated as having level I or II evidence, 11 had a calculated LR for a positive result of > 5, 9 had LRs of > 10, and 3 had LRs of > 20 (Table 4). The type of signal used to detect events had no obvious impact on the reported results for specificity and sensitivity. Data loss was not reported in a number of studies. When reported, data loss ranged from 0% to about 11%. The majority of studies had 30 to 60% of patients testing positive on the portable monitor. Those studies reportingLRs of around ≥ 10 had a reasonably small rate of false-positive results (range, 0 to 12%), while studies with lower LRs for a positive result had a much higher percentage of false-positive results (range, 3 to 37%).

4.1.2.3.2. Home-unattended

There were 8 studies reporting sensitivity/specificity data on type 4 portable monitoring performed in the home setting, four of which were rated as having level I or II evidence,[33,40,41,53] with the rest rated as having level IV evidence. All eight studies utilized oxygen saturation monitoring (seven studies used it as the primary criterion, and 1 study used it as a secondary criterion combined with flow measured by thermistor). Three of the four studies with evidence level I/II[33,39,41] (oxygen saturation) had LRs of > 10, and one study[53] had a very low LR for a positive result. Two of the studies with level IV evidence also reported high specificities/LRs. The prevalence in most of these studies was between 40% and 70%. The percentage of patients testing positive on the portable monitor among the studies with a high calculated LR ranged from 14 to 42%, with the range of false-positive results in the high LR studies being 0 to 12%. The percentage range of positive results and the percentage range of false-positive results in the three studies with a calculated low LR were 47 to 73% and 10 to 41%, respectively.

There were several different approaches for detecting and scoring events, including desaturation index, desaturations plus heart rate change, a delta index and cumulative time below a designated saturation, and visual qualitative assessment. Some portable monitors used more than one approach. Scoring was performed by computer without editing in two studies, by manual scoring in three studies, by both in two studies, and was not reported in the final study (Table 1). There was no consistent pattern indicating that computer scoring or manual scoring was superior, although one study (using thermistor plus oxygen desaturation) that compared them directly found that manual scoring was better.[33]

4.1.3. Evidence that a single portable monitor can be used to both reduce and increase the probability that a patient has an abnormal AHI

If a portable monitor is going to be used to both reduce (given a negative result) and increase (given a positive result) the probability that a patient has

sleep apnea, it has to have both a high sensitivity/low LR for a negative result and a high specificity/high LR for a positive result. If a monitor is able to achieve this at a single threshold, then patients will have either positive or negative results, and if the test has excellent operating characteristics there will be few false-positive or false-negative results. However, if the low LR is at one threshold and the high LR is at a different threshold, then some patients will not be classified as having a positive or negative result by the portable monitor. The information on which monitors had both a high positive LR and a low negative LR is summarized in Tables 5 and 8. The percentage of patients who were not classified by the monitor is also included in Table 5. This number was 0 if the monitor used a single threshold value to define positive and negative results. The percentage of patients who had a false result (*ie*, a false-positive or false-negative result) is also reported in Table 5.

Twenty-eight study groups from 27 articles presented data that indicated that the portable monitor in the study could both reduce (low LR of a negative test) and increase (high LR of a positive test) the probability of sleep apnea in the population studied.

4.1.3.1. Type 2 monitors

Only two studies of type 2 monitors (both rated as having level IV evidence) reported results of high and low LRs.

4.1.3.1.1. Sleep laboratory-attended

A single study[32] (rated as having evidence level IV/quality rating d) with a small number of patients (40 patients) reported a very high sensitivity and specificity at the same threshold, resulting in a very small percentage of patients being misclassified as having either a false-negative or a false-positive result (identified in Table 5 as the percentage of patients with a false result).

4.1.3.1.2. Home-unattended

One study[20] (rated as having evidence level IV/quality rating b) showed a modest sensitivity but a very high specificity, with 10% of patients being misclassified. This study also had a high rate of data loss (20%).

4.1.3.2. Type 3 monitors

4.1.3.2.1. Sleep laboratory-attended

Of nine studies using type 3 monitors in a sleep laboratory setting, all of them reported data that produced both high positive and low negative LRs. As previously described, eight of these studies[18,21–23,27–29,31] were rated as having either level I or II evidence. Only one study[29] had patients who were not classified as having either positive or negative results, and most studies had a misclassification rate (*ie*, a false result) of only approximately 5%.

4.1.3.2.2. Home-unattended

Two of the four home studies (both rated as having evidence level II) using type 3 monitors[26,29] reported data that provided both high positive and low negative LRs. They both used flow measured by thermistor and an optional oxygen desaturation criterion for detecting events. Both studies used different thresholds to produce the low and high LRs, with the result that 22% of patients in one study[29] and 37% of patients in the other[26] had an intermediate result and were not classified as having, or not having, sleep apnea. The misclassification rates were 16% and 5%, respectively, for the two studies.

As stated in the preceding sections, the majority of the evidence from attended monitoring indicates that type 3 monitors using flow measured by thermistor can substantially reduce and increase the probability of sleep apnea in the vast majority of patients. The limited data from the home setting indicate that this approach requires additional evaluation in that setting.

4.1.3.3. Type 4 monitors

4.1.3.3.1. Sleep laboratory-attended

Of the 25 studies published on type 4 monitors in a sleep laboratory setting that reported sensitivity and specificity, 15 of them (level I evidence, 2 studies; level II evidence, 7 studies; level IV evidence, 6 studies) reported data that produced both high and low LRs. Ten of these studies utilized different thresholds to achieve both low and high LRs (Table 5). Some of these studies[13,34,36] reported a low percentage of patients who were not classified as having positive or negative results, but most[12,33,35,40,43,49,50,51,57] had substantial proportions of patients who were not classified as having positive or negative results. Misclassification rates tended to be low (*ie*, $< 7\%$). Of the five laboratory studies that used a single portable monitoring threshold, four[13,14,17,34] had $< 7\%$ false results.

There was no obvious pattern suggesting that different approaches for detecting breathing events improved the LRs, the percentage of patients not classified, or the misclassification rate. Studies rated as having level I, II, or IV evidence all reported variable rates of unclassified and misclassified patients, with no obvious differences between different techniques used for event detection.

4.1.3.3.2. Home-unattended

A single study[40] (that used oxygen saturation) on a portable monitor studied in the home setting reported data that allowed the monitor to classify some patients as either having or not having sleep apnea. There was

a high percentage of patients not classified (50%), but there was a low misclassification rate (4%).

Overall, most of the studies (19 of 27 studies) reporting data that produced both low and high LRs were rated as good quality (*ie*, evidence level I/II, with most quality indicators met). The monitor type or primary signal measured did not appear to influence the results. The prevalence of sleep apnea was high (*ie*, $> 40\%$) in the majority of studies, but not all. All studies had a wide range of sleep apnea severity represented in the population studied. Many of the studies utilized manual scoring, but several did not as the portable monitor was not designed to allow for it (Table 1).

4.2. Secondary Questions

4.2.1. Reproducibility of portable monitoring

The reproducibility of portable monitor results can be viewed from several perspectives. The first is the reproducibility of the human scoring of events. When monitors are scored or edited manually, there should be some indication of how likely it is that a single reviewer can get the same answer on separate occasions when scoring the same records (*ie*, intrarater reliability) as well as how likely it is that two separate reviewers will get the same answer when scoring the same records (*ie*, interrater reliability). Of the 51 studies that were reviewed, 19 reported manual scoring exclusively, and 10 incorporated some type of manual editing of automated scoring or combined manual and automated scoring (Table 1). None of these 29 studies reported interrater or intrarater reliability.

Another important source of variability in the results of monitoring is that seen from night to night. This could be random variability or due to a first-night effect similar to what has been described with polysomnography.[81] Only two studies reported results on night-to-night variability. Both studies were conducted using portable monitoring that was performed unattended at home, and both studies were categorized as having evidence level II and quality rating a, and both utilized thermistor-measured flow plus oxygen desaturation as a criterion for breathing events. The first study[18] reported data on 32 subjects who were studied for 2 nights at home and found no differences in the degree of desaturation or the duration of events. A correlation coefficient of 0.94 was reported between the RDI measured the first night and the RDI measured the second night. A Bland-Altman analysis was not reported. There was no statistical difference in the mean RDI between the two nights, suggesting that there was not a first-night effect. Using an RDI of > 10 to define an *abnormal study*, they found that only one subject was reclassified as having abnormal results on the second study after initially being classified as having normal results on the first study. The second study[33] reported the difference in portable monitoring results between the first night in the sleep laboratory and a second night at home. Twelve of 53 patients had a difference in RDI of ≥ 10 between the two nights.

The third check on the repeatability of a monitor's results would be independent validation between different groups of investigators using nearly identical study protocols. Although several studies used similar monitors, differences in study protocols would preclude any conclusion that a single monitor had been properly evaluated for consistency in different sleep centers.

At this time, there have been a very small amount of data reported on the reproducibility of portable monitoring results. It is an obvious area for further research.

4.2.2. Cost-benefit analyses of portable monitors

There were a limited number of studies that included commentary and data on cost comparison between portable monitoring and polysomnography. Since the focus of these articles was on a diagnostic comparison with polysomnography, formal cost-benefit analysis methods were not used. Reports variably mentioned effectiveness, benefits, or costs of portable monitoring alone or examined the potential cost savings if polysomnography had been avoided. When conducting the latter analysis, most investigators assumed the necessity of a dual-night monitoring (*ie*, diagnostic and therapeutic) format and did not consider the reduced cost for a single, split-night polysomnogram.

When performed, the analysis of costs was limited to the obvious direct costs of performing portable monitoring and polysomnography. The broader impact on society, such as access to sleep apnea diagnostic testing, the indirect costs of not diagnosing and treating patients (*eg*, motor vehicle accidents, industrial accidents, and lost time from work) that could occur either by missing the diagnosis with a false-negative portable monitor test result or because the alternative polysomnogram was not accessible, was not considered.

Many of the studies make inferences about cost savings that would occur based on favorable conclusions regarding portable monitor sensitivity and specificity. This section will focus on those studies in which numbers are provided in terms of actual dollars, time, or the number of additional tests avoided.

4.2.2.1. Type 2 monitors

Of the four studies on type 2 portable monitors, only two (both having level IV evidence and a quality rating of a or b) commented on cost. One study[19]

mentioned a projected cost reduction based on the assumption that overnight staff and the facility cost could be eliminated with a portable monitor, but no data were provided. The second study[20] indicated that the portable monitoring was half the cost of laboratory-based polysomnography (patients who preferred laboratory-based polysomnography, 48%; patients who preferred the home-based study, 28%; and patients who had no preference, 24%). The 20% failure rate with this monitor, which was set up in the home, was not accounted for.

4.2.2.2. Type 3 monitors

Twelve type 3 portable monitor studies were reviewed, however, 6 studies did not mention cost, and 3 studies[24,29,60] simply presumed that portable monitoring would be associated with reduced cost because less technician time was involved (*ie*, set-up and/or scoring). One sleep laboratory-based assessment[23] (level II evidence and quality rating a) in 104 patients compared the actual setup and data verification time for portable monitoring and standard polysomnography, and reported setup times of 20 to 25 min vs 60 to 75 min, respectively, and data management times of 45 to 90 min vs 180 min, respectively.

There were two studies of portable monitoring performed in the home that reported costs. One study[26] (level II evidence and quality rating a) from Spain (89 patients) calculated (1) the direct costs of portable monitoring testing (accounting for the single use of the equipment, home setup with and without a technician, and monitor failure requiring repetition of studies) and (2) the indirect costs, including travel and repair. The effectiveness of the portable monitor was 92% (false-negative result rate at a threshold AHI of 10, 8%) with technician home setup and 80% with patient home setup (with added 12% failure rate). The reported cost of portable monitoring was approximately one third that of polysomnography, whether or not the technician did the setup.[26] The second study[30] (level IV evidence and quality rating b), from the United Kingdom (123 of 150 patients had interpretable data) reported that the portable monitor cost 21.9% of their sleep laboratory-based polysomnography fee. The wait for investigation with sleep laboratory-based polysomnography vs portable monitoring was reduced from a median of 47 days to 18 days ($p < 0.001$). When looking at the entire population of subjects, accounting for patients who could avoid polysomnography as well as those having to undergo both tests, the cost saving was small at £3.17 per subject.

4.2.2.3. Type 4 monitors

Analysis is complicated by the variability of the equipment used and the parameters monitored, as well as the chosen focus of cost or benefits reported in each article. Of 35 type 4 monitor studies, 8 studies provided some comment on cost-benefit-related issues, of which 5 were studies of attended portable monitoring, 2 were studies of unattended monitoring in-home, and 1 was a study using a mixed-population methodology.

A level IV, quality rating c, unattended home oximetry study from the United Kingdom[56] used the British Thoracic Society criteria of ≥ 15 desaturations ($\geq$4%) per hour spent in bed. Since their sensitivity was low (31%) and the specificity high (100%), the assumption was that only symptomatic patients with negative home oximetry study results needed to undergo follow-up polysomnography. Screening all 69 patients with oximetry (estimated cost, £10) would have avoided 10 polysomnograms (estimated cost, £500) for a savings of £62 per patient. The other unattended home oximetry study from Canada[53] (level I evidence and quality rating b) reported a high sensitivity (98%) but poor specificity (48%) and concluded that 62 patients with negative oximetry study results (26%) of a total of 240 patients would not have needed polysomnography. False-negative results were found in 2 of the 62 oximetry-negative patients, and 18 studies were repeated in patients reporting poor sleep.

Two of the six studies that used oximetry in an attended setting made conclusions about potentially eliminating the need for polysomnography but made different assumptions. The investigation from the United States[35] (level II evidence and quality rating d) used a multivariable apnea prediction questionnaire with oximetry and concluded that their combined algorithm in a high-risk population reduced the number of polysomnography studies by only 8 to 12% in patients with negative oximetry study results using an RDI threshold of 30. They added, however, that the algorithm also provided sleep specialists with a tool to identify high-risk patients that should be studied quickly. The other simultaneous laboratory-based oximetry article from Spain[45] (level I evidence and quality rating b) suggested that symptomatic patients with negative oximetry study results should undergo a polysomnogram in addition to those with positive oximetry study results who were also known to have abnormal pulmonary function test results (presumably more complex disease but details not specified). Using a 4% desaturation criterion and examining RDI thresholds of 5, 10, or 15 suggested that if oximetry was performed initially, polysomnography could have been avoided in 140 (51%), 119 (43%), or 105 (38%), respectively, of the 275 patients analyzed.

There were two studies of attended monitoring using nasal pressure to detect airflow that reported

on costs. A study from the United Kingdom[14] (evidence level IV and quality rating b) with 31 patients computed the actual dollar figures, including the depreciation costs of equipment analyzed for 200 studies per bed per year, at £14 for the portable monitor vs £126 for polysomnography. Assuming those patients with an AHI of ≥ 20 also needed continuous positive airway pressure (CPAP) titration, the cost per patient was estimated at £35 for portable monitoring compared to £154 for polysomnography. Another study from France[12] (evidence level II and quality rating a) that used the same portable monitor did not directly assess financial factors but estimated the reduction in the number of polysomnography studies if portable monitoring were used first. When a clinical algorithm that rated a patient's pretest probability of having an AHI of ≥ 15 indicated a probability of $> 80\%$ (46 patients), only 5 patients who had a negative portable monitor study result would have required polysomnography, saving 41 sleep studies. From their entire population of 95 patients, only 25 patients with a negative portable monitor study result would have required subsequent polysomnography, saving a total of 70 polysomnograms.

Two other studies used unique equipment with additional channels added to oximetry, including heart rate and snoring in one study (evidence level II and quality rating a) and thermistor-derived nasal airflow in the other. The first of these studies[43] was performed in Spain and used an attended sleep-laboratory study design in 150 patients. The algorithm for that study included the use of portable monitoring in all patients ($155 [in US dollars]) utilizing polysomnography ($546) in those who had doubtful results (25%) and in those who required repeat testing (10%). They concluded that with this approach in their population of patients with clinically suspected sleep apnea there would be a net cost reduction of 44%. The final study[33] from Canada (evidence level II and quality rating a) considered an algorithm in which polysomnography was required only if the portable monitor showed an RDI between 2 and 20. They attributed 4 h of technician time to full polysomnography. The investigators concluded that an initial portable monitor study with this protocol and an 8% data loss would exclude the need for 42 polysomnograms of 68 patients for a net reduction in total technician time of 2 h per patient (95% confidence interval, 1.6 to 2.5 h).

4.2.2.4. Summary

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

4.2.3. Failure rate of portable monitors

4.2.3.1. Type 2 monitors

Data loss was reported for three of the four studies. One study[19] performed at home reported no lost studies in 77 patients but reported a loss of oxygen saturation data in 4% of patients. The other unattended home study[20] reported a 20% loss of data.

4.2.3.2. Type 3 monitors

Six studies explicitly reported the percentages of portable monitoring studies that did not collect adequate data. These ranged from 3%[24,29] to a high of 18%[30] in studies performed unattended at home. Of the two studies performed in the sleep laboratory, data loss was reported in 3.3% of patients in one study[21] and 9% in the other.[18]

4.2.3.3. Type 4 monitors

Fourteen studies explicitly reported the percentages of portable monitoring studies that did not collect adequate data. One study[33] reported the data loss for patients at home and in the laboratory at approximately 8% in both locations using a thermistor to measure flow and an oxygen saturation monitor. There were three other unattended home studies[41,42,53] that reported data loss of 7 to 10%. All of these studies used oxygen saturation as the signal for defining the occurrence of events. Of the 10 studies performed in an attended sleep laboratory setting, data loss varied between 2%[36] and 14%[48] for monitors using oxygen saturation, and between 10% and 14% for four monitors that used nasal pressure to detect flow.[10,13,15,17] One monitor that used heart rate variation had a signal failure rate of 5 to 11%.[52,58]

4.2.4. Types of patient populations studied: generalizability of findings

4.2.4.1. Disease severity

4.2.4.1.1. Type 2 monitors

Three of the four type 2 monitor studies[20,25,51] reported data on the distribution of AHI values in the population of patients studied. These studies included patients with a wide range of AHI values (as

indicated by a mean AHI of > 20 events per hour) and included patients who did not have sleep apnea (*ie*, AHI, < 5) as well as patients with more severe sleep apnea (*ie*, AHI, > 40).

4.2.4.1.2. Type 3 monitors

All 12 type 3 monitor studies included patients who did not have sleep apnea (*ie*, AHI, < 5). Eleven of the 12 studies included patients with severe sleep apnea, and while the range of AHI values was not reported in the 12th study,[22] 3 studies[23,27,31] reported a mean AHI of < 20. One study[22] did not report the AHI distribution in their study patients.

4.2.4.1.3. Type 4 monitors

Thirty-two of the 35 type 4 monitor studies included a range of patients from those who had no sleep apnea to those who had severe sleep apnea. Of these 32 studies, 6 reported a lower mean AHI (*ie*, < 20)[13,16,17,33,47,57] than the other studies. The range of AHI values was not reported in three studies.[40,48,54]

4.2.4.1.4. Summary

The operating characteristics of a portable monitor should be tested in patients with and without sleep apnea, and in patients with a wide range of AHI values. The majority of studies included patients with a wide range of AHI values, as indicated by both the presence of severe AHI levels and a mean AHI of > 20. This is perhaps not surprising given that the vast majority of studies were performed in patients who were referred to sleep clinics or sleep laboratories for polysomnography for the diagnosis of sleep apnea (see section 4.2.4.4). Because the mean AHI in most studies was in the moderate-to-severe range of severity (*ie*, AHI > 20), these findings may not be generalizable to populations with a lower pretest probability (*eg*, general population, women, and primary care population).

4.2.4.2. Comorbid conditions

4.2.4.2.1. Type 2 monitors

Only one study[20] reported whether patients with comorbid conditions (*eg*, chronic bronchitis [no criteria given]) were included in the study. One study[25] reported other sleep diagnoses that were made after all investigations were included.

One study noted that the patients studied included those referred for "sleep-related complaints" (*ie*, not just sleep apnea) but did not give a breakdown of all sleep diagnoses made. No mention of other sleep disorders was made in the remaining two articles.

4.2.4.2.2. Type 3 monitors

Of the 12 studies using type 3 monitors, 9 did not make note of any comorbid conditions that the patients may or may not have had, including lung disease. One study[18] included patients with obstructive and restrictive lung diseases. Two studies[31,60] specifically excluded patients with "significant" lung disease, although no definitive criteria were provided. One study[31] excluded patients with greater than class 2 New York Heart Association congestive heart failure.

Three studies[24,28,30] specifically noted that patients who were suspected of having other sleep disorders were excluded from their study protocol. One study[23] stated that patients with other sleep disorders were included, and final diagnoses were provided. All other studies included only subjects who were suspected of having sleep apnea.

4.2.4.2.3. Type 4 monitors

Of the 35 studies using type 4 monitors, 23 did not make note of any comorbid conditions that patients may have had, including lung disease. Three studies specifically mentioned that patients with COPD[41,43] or COPD and restrictive lung disease[52] were included in the study group. Two studies[46,50] used patients with lung disease as a comparison group. Several studies specifically excluded patients with significant lung disease, as defined by an FEV_1 of < 50% predicted,[36] daytime hypoxemia (*ie*, oxygen saturation of < 90%[40,56] or < 7.3 kPa[36]), or the use of supplemental oxygen.[35,49] One study[13] excluded patients with significant COPD and awake hypoxemia but did not give specific cutoffs. One study[58] excluded patients with congestive heart failure or insulin-dependent diabetes mellitus but not lung disease.

Only three studies[16,40,54] made note of other sleep-related diagnoses that were made after all investigations had been performed. No study specifically mentioned excluding patients with other sleep disorders.

4.2.4.2.4. Summary

Many patients who present to a sleep center/laboratory for the evaluation of sleep apnea have received other pulmonary diagnoses including asthma, COPD, and obesity-hypoventilation syndrome. COPD and obesity-hypoventilation syndrome are associated with nocturnal oxyhemoglobin desaturation. The presence of these disorders could influence the performance of a portable monitor, particularly those in which oximetry is the primary parameter measured. Congestive heart failure is also a common diagnosis in patients who are referred for sleep apnea evaluation and could influence portable monitoring results if the predominant respiratory abnormality during sleep is Cheyne-Stokes breathing or central sleep apnea. The majority of studies

included in this systematic review did not indicate whether the subjects had significant comorbid conditions, such as lung and heart disease, that could influence the performance and outcomes of portable monitoring. Therefore, at this time, the findings of this review should be applied only to patients without significant pulmonary or cardiac comorbidity.

Furthermore, sleep centers evaluate a wide variety of sleep complaints such as narcolepsy and periodic limb movement disorder. The presence of patients with these disorders in a portable monitoring study for sleep apnea evaluation would reduce the prevalence of sleep apnea in the study population. The majority of studies included only subjects in whom there was a clinical suspicion of sleep apnea and did not provide alternative sleep diagnoses for those subjects who were found not to have sleep apnea. Therefore, the findings from this review can only be generalized to patients with a clinical suspicion of sleep apnea and not necessarily to all patients with other possible sleep disorders who are referred to a sleep center.

4.2.4.3. Gender/race

4.2.4.3.1. Type 2 monitors

Three studies reported the ratio of men to women in their patient population, with men accounting for > 70% of the sample in two of them. None of the studies reported whether there were nonwhite subjects studied.

4.2.4.3.2. Type 3 monitors

Eleven of 12 studies reported the ratio of men to women in their patient population, with men accounting for > 70% of the sample in 10 of these studies and 56% in the other study.[27] None of the studies reported whether they included nonwhite subjects.

4.2.4.3.3. Type 4 monitors

Of the 35 type 4 monitor studies, 27 reported the number of men and women studied. In 24 of these reports, men made up > 70% of the subjects studied. One study[35] reported a breakdown of their study population by race (white, 72%; black, 24%; other, 4%).

4.2.4.3.4. Summary

Women have been shown to have a lower prevalence of sleep apnea,[2] and those with the disorder predominately have hypopneas.[82] One study[83] has suggested that African-Americans have a prevalence of sleep apnea similar to that of whites, but that they have a higher AHI after correction for body mass index and age. Typical of most referral populations to sleep centers, the majority of studies in this review included predominately male subjects. Generalizing these findings to other populations in which there are a greater percentage of female subjects may not be appropriate. In addition, because only a single study indicated the inclusion of nonwhite subjects, it is unclear whether these findings can be generalized to nonwhite patients.

There are no data in the literature to suggest that the measurement of respiratory-related events during sleep is influenced by gender or race. Research is required to delineate the utility of portable monitors in populations of nonwhites and in those populations with a greater percentage of women.

4.2.4.4. Population: patient recruitment

4.2.4.4.1. Type 2 monitors

Only one study[19] stated that the study recruitment process began in a sleep clinic. However, portable monitoring was performed only in those patients in whom polysomnography was deemed to be necessary. The other three studies were performed in patients who had been referred to a sleep laboratory and did not give specific criteria for how these patients had been referred.

4.2.4.4.2. Type 3 monitors

Of the 12 studies using type 3 portable monitors, only 1 study[27] was performed in a subset of subjects who had been included in a population study of sleep apnea. Patients were recruited from a sleep clinic in six studies.[21–23,28,30,31] However, in three of these studies,[23,28,30] portable monitoring was performed only in patients in whom polysomnography had been deemed to be necessary. The remaining six studies were performed in a sleep laboratory setting, and the criteria for referring the patients to the laboratory for study were not included.

4.2.4.4.3. Type 4 monitors

Of the 35 type 4 portable monitoring studies, 22 included only patients suspected of sleep apnea who had been referred to a sleep laboratory for polysomnography. No study gave the specific criteria or reasons why these patients had been referred to the sleep laboratory for study. However, three studies[39,45,49] did note that patients had been referred to the laboratory by a variety of physician specialists, including generalists, hospitalists, pulmonologists, and otolaryngologists.

Thirteen studies[34–36,40–42,45,51–53,56–58] specifically mentioned that the patient recruitment began in a sleep clinic. However, only seven studies[36,38,40,45,49,53,56] stated that study patients were a consecutive or random sample of all patients who had been referred to the clinic for sleep apnea evaluation. These seven studies performed diagnostic testing on all patients who had been referred for sleep apnea evaluation to the clinic, not just on patients suspected of having sleep apnea after a physician's evaluation.

4.2.4.4.4. Summary

The population in which a diagnostic test is being performed is critical to the interpretation of the reported operating characteristics of the test such as sensitivities, specificities, and LRs. For sleep apnea, the pretest probability is low in the general population, higher in a sleep clinic population, and highest in a sleep laboratory population. In this review, the majority of studies (30 of 51 studies) were performed in patients who had been referred to a sleep laboratory for suspicion of sleep apnea. The exact criteria used by the referral physicians to make the decision to order polysomnography were not provided. Using a population of patients who have been preselected by some group of specialist physicians to study a portable monitor makes it difficult to know whether the results can be generalized to patients who have not been assessed by those physicians. Because there was only one study[27] performed in a general patient population, there was not enough information available to predict whether the findings from the other 50 studies are applicable to non-sleep clinic/laboratory populations.

4.3. Summary

4.3.1. Overview and study limitations

In systematic reviews of topic areas such as the use of portable monitoring for the diagnosis of sleep apnea, the AHRQ suggests basing recommendations on the quantity, quality, and consistency of the research findings.[66] The ATS/ACCP/AASM working group that reviewed the evidence thought that a formal meta-analysis of the research on this subject with summary ROC curves was not possible because of the marked heterogeneity of methods, definitions, and monitor types/signals. Fifty-one published studies were reviewed. The number of studies on a single monitor ranged from one to eight. Within the category of oximetry, variability existed, because in some studies oximetry was used alone while in others it was used with another sensor (*eg*, heart rate or snoring). As well, each study used a different method of scoring the signal. Therefore, it is not possible to justify a recommendation on any single monitor because the quantity and consistency of results cannot be adequately assessed at this time.

The quality of the studies on portable monitors for diagnosing sleep apnea varied, although the majority were of a high evidence level and quality rating (*ie*, evidence level I and II, and quality rating a or b). There were few evidence level I studies performed, because most authors did not use consecutive patients who were presenting to a sleep center. The majority of studies used patients who had been referred to a sleep laboratory, which increased the chance that there was selection bias. Most studies having evidence level II met the majority of other quality indicators (*ie*, had a quality rating of a or b). Common reasons for having a lower quality rating included poor characterization of the equipment used, lack of definitions of respiratory events, lack of description of blinding, and the inability to perform a manual review of the automated analysis.

Several limitations of the published studies may prevent wide applicability of the findings, as follows:

1. All studies were performed in sleep centers/laboratories that had a high pretest probability (*ie*, prevalence) of sleep apnea. In this setting, the number of false-negative results increases and reduces the negative predictive value of portable monitors.
2. The studies primarily include men without significant comorbid illness, so their applicability to women and to patients with congestive heart failure or COPD is unknown.
3. The number of patients in the reviewed studies varied widely (range, 20 to 359), which directly impacts the precision of the result and the confidence that the results are reproducible.
4. The majority of studies report comparisons of attended portable monitors with polysomnography, so the effectiveness of portable monitors in a home setting is not as well-established.
5. Using time in bed as the reference for type 3 and 4 portable monitors may lead to a lower RDI than would be expected for the AHI ascertained by polysomnography.
6. Using dual-RDI thresholds (a low one to exclude disease and a higher one to confirm disease) leads to the presence of a number of patients who cannot be classified as having either a high or low likelihood of having an abnormal AHI.
7. The oximeter sampling rate may not be sufficient to capture all of the arterial desaturations.
8. The design of the unattended portable monitoring studies did not allow the results of the comparison with polysomnography to be adjusted for the effects of night-to-night variability, and differences in body position (*eg*, supine vs lateral) and sleep stages (*eg*, rapid eye movement [REM] sleep vs non-REM sleep).
9. The findings of any one study are integrally linked and therefore limited by the design of the technology at the time the protocol was performed. Since the results were published, many monitors have undergone modifications that would likely impact on the results found in future studies.

10. Type 3 and 4 portable monitors do not detect cortical arousals. In as much as this is an important part of event definition, this could be viewed as a limitation of these types of monitors.

4.3.2. Primary questions and monitor types

In this section, a summary of the studies by monitor types that address the three primary questions are discussed. Because most monitors had only one or two studies published about them, and those were the ones that had more variable research designs, the comments are limited to the number (*ie*, quantity) of studies having a high level of evidence and a high quality rating (*ie*, evidence level I or II, and quality rating of a or b) that specifically answered the question compared to the total number of high-quality studies (*ie*, consistency). The number of studies with low false-negative results or false-positive results (defined as $\leq 10\%$) also will be discussed in the context of utility.

4.3.2.1. Type 2 monitors

Type 2 monitors allow for sleep staging and should give the best agreement with polysomnography because an AHI can be calculated. However, there were only four published studies[19,20,25,32] utilizing type 2 monitors, and one of those studies[19] did not provide sensitivity/specificity data. Two of the remaining three studies had lower grade evidence (level IV evidence). Because of the small quantity and low quality of these studies, summary statements about consistency cannot be made, and these monitors are not discussed in the sections below.

Synopsis: *Convincing evidence indicating that type 2 monitors could be used in either an attended or an unattended setting is lacking.*

4.3.2.2. Evidence that portable monitors can be used to decrease the probability that a patient has an abnormal AHI (LR, < 0.2)

4.3.2.2.1. Type 3 monitors

Twelve studies reported 13 comparisons between portable monitoring and polysomnography. All studies used a thermistor to define apneas and hypopneas on polysomnography (most also included an oxygen desaturation criterion), and respiratory events on portable monitoring (one study[60] failed to report how events were defined) [Table 1]. Eight of nine attended monitor studies were of higher evidence level and quality rating (three studies had level I evidence and a quality rating of a), and all had a low LR (*ie*, < 0.2) [Table 3]. Seven of the eight studies[21–23,27–29,31] had a low percentage of false-negative results. In contrast, two of the four unattended studies[26,29] had a higher level of evidence and higher quality rating (both had level II evidence, and a quality rating of a or b). Both studies had low LRs but a relatively high percentage of false-negative results.

Synopsis: *Several studies with a high level of evidence and high quality consistently show that some monitors have utility to decrease the probability of sleep apnea in an attended setting. In an unattended setting, the results should be considered preliminary and suggest that these monitors may be useful, but that their actual utility requires additional study.*

4.3.2.2.2. Type 4 monitors

There were seven comparisons (six studies[33,34,39,43,44,57]) that combined oximetry with at least one other parameter, such as heart rate or snoring, to score respiratory disturbances. Five studies[33,34,43,44,57] were performed in an attended setting, and all had level II evidence, a quality rating of a, and a low LR. One study[57] had a high rate of false-negative results. One study[33] of the two unattended studies[33,39] had a higher level of evidence (IIa), but both had a relatively high rate of false-negative results.

There were 19 comparisons (18 studies) evaluating the use of oximetry alone as a reference. Thirteen of these comparisons were attended studies, with 6 having a higher level evidence and a higher quality rating,[36,37,45,52,54,55] and 4 having both a low LR and an acceptable level of false-negative results.[36,37,52,55] Of the six unattended home studies, only one was of a high level of evidence and high quality.[53] This study had both a low LR and a low percentage of false-negative results.

There were seven studies of nasal pressure, all of which occurred in an attended setting. Only three studies[11–13] had a higher level of evidence and a higher quality rating. All three studies had low LRs, and two of them[11,13] had a low percentage of false-negative results.

Synopsis: *Oximetry alone can reduce the probability of sleep apnea both in an attended and unattended setting. However, in the latter situation the results should be considered preliminary. The addition of a second signal showed results that were similar to those using oximetry alone (although there were fewer studies evaluated), and similar conclusions can be drawn. Nasal pressure may be useful in an attended setting, but no conclusions can be made about its use in an unattended setting.*

4.3.2.3. Evidence that portable monitors can be used to increase the probability that a patient has an abnormal AHI (LR, > 5)

4.3.2.3.1. Type 3 monitors

Eight of nine studies in an attended setting were of higher quality, and all had a high LR. However, the LRs ranged from 6.4 to infinity (Table 4). Five of

the eight studies[18,21,22,28,31] had a low percentage (*ie*, < 10%) of false-positive results. Two of the four studies[26,29] performed in an unattended setting were of higher quality, both of which had marginally high LRs (range, 5 to 9), with one study[29] having a high percentage of false-positive results.

Synopsis: *There are several high-quality studies showing that type 3 monitors can increase the probability of sleep apnea in an attended setting. The data supporting the usefulness and utility for increasing the probability of sleep apnea in the unattended setting are limited.*

4.3.2.3.2. Type 4 monitors

Of the six comparisons made in the five studies[33,34,43,44,57] of type 4 monitors using oximetry and a second parameter having higher evidence levels and higher quality, five comparisons (four studies[33,34,43,57]) had a high LR (range, 10.3 to 45). Two studies[33,44] had a high percentage (*ie*, > 10%) of false-positive results.

Of the eight studies[35–37,45,48,52,54,55] of oximetry alone in the attended setting with higher levels of evidence and higher quality, five[35,36,45,48,54] had high LRs (range, 8.2[36] to 18[35]). One of the five studies[36] had a high percentage of false-positive results. Of the three studies[40,51,53] with a higher level of evidence and higher quality performed in the unattended setting, two[40,41] had a high LR with acceptable levels of false-positive results.

Only two studies[12,13] of the three[11–13] performed on nasal pressure that were of a higher level of evidence and higher quality showed high LRs, but only one[12] had a low rate of false-positive results in the attended setting. There were no studies performed in the unattended setting.

Synopsis: *Oximetry alone can increase the probability of sleep apnea both in an attended and unattended setting. However, in the latter situation the utility appears to be less compared with that in the attended setting. The addition of a second signal showed results similar to those using oximetry alone. However, the evidence is lacking to suggest that this type of signal combination can be used in an unattended setting. The results for nasal pressure in the attended setting should be considered preliminary but suggest that the utility of this approach is still in question. No conclusions can be made about its use in an unattended setting.*

4.3.2.4. Evidence that a single portable monitor can be used to both reduce and increase the probability that a patient has an abnormal AHI (negative LR, ≤ 0.2; positive LR, > 5)

Twenty-eight comparisons from 27 studies reported data indicating that the portable monitor in the study could both reduce the probability of sleep apnea in the population studied (*ie*, low LR for a negative test result) and increase the probability of sleep apnea in the population studied (*ie*, high LR for a positive test result) [Table 5]. The majority of these studies had higher levels of evidence and a higher quality rating (level I or II evidence and quality rating of a or b, 18 of 27 studies). The portable monitor sensor configurations varied widely, as did the number of patients studied, the severity of sleep apnea in the patient populations, and the definitions of sleep apnea used to define abnormality. The prevalence of sleep apnea was high in the majority of studies, but not in all. Some studies relied on automated scoring, others used manual scoring, while others utilized both (Table 1).

Sensitivity and specificity were often calculated for multiple levels of RDI, resulting in a range of results that would either exclude or confirm the diagnosis of sleep apnea. In 15 of these studies, the best sensitivity and specificity were at the same RDI level, while 13 studies found that the best sensitivity and specificity were at different RDI levels. In the latter case, some patients will have a nondiagnostic result (next-to-last column of Table 5), which will be used as an indicator of utility in the following sections.

4.3.2.4.1. Type 3 monitors

In the attended setting, all eight studies of type 3 monitors having a high evidence level and high quality rating had both high and low LRs, indicating that the portable monitor could be used to both confirm and exclude the diagnosis of sleep apnea. One study[29] used multiple cutoffs and had a high rate of patients without a negative or positive test result. Two[26,29] of four unattended studies of type 3 monitors having high evidence levels and high quality ratings had both high and low LRs. Both studies used different RDI levels, and both had a high rate of patients without a negative or positive test result.

Synopsis: *Type 3 monitors have utility to both reduce and increase the probability that a patient may have sleep apnea in the attended setting. The utility in the unattended setting is not as well-established.*

4.3.2.4.2. Type 4 monitors

There were four studies[33,34,43,57] of type 4 monitors with oximetry and at least one other sensor, all of which were performed in the attended setting, that had both high and low LRs. All were of higher quality. Three studies[33,43,57] used multiple cutoffs to achieve the high and low LRs, and had a high rate of patients who did not have a diagnostic result on the portable monitor study.

There was one unattended home study[40] of a portable monitor using oximetry only that had both

high and low LRs. This study had a high rate of nondiagnostic results. In addition, three of nine studies[34–36] using oximetry alone having a higher evidence level and higher quality rating in the attended setting that had both high and low LRs. Two of these studies[35,36] had a high rate of patients without a diagnostic test result. Two higher quality studies of nasal pressure (both performed in the attended setting) showed both high and low LRs, one of which had a high nondiagnostic rate.

Synopsis: *The utility of using oximetry alone to both increase and reduce the probability of sleep apnea is not well-established in the attended or unattended setting, with or without an additional channel. Results should be considered preliminary at this point. The utility of nasal pressure as a signal for increasing and decreasing the probability of sleep apnea has not been established in the attended or unattended setting.*

4.3.3. Summary discussion

Clinicians interested in using portable monitors have different reasons for doing so. Some wish to use them as screening devices to exclude the presence of the disease, while others wish to use them to decide whether to initiate treatment. Several monitors show promise for excluding disease (Table 3), others for confirming disease (Table 4), and some for doing both (Table 5). Clinicians interested in using portable monitoring for investigating patients with suspected sleep apnea need to review carefully what they want the monitor to do, the key signals they believe it should record, the quality of the research that has been published, and whether the study setting/patient population is similar enough to their own that it makes sense to accept the published results. At this time, since it is not possible to base a recommendation on the quantity and consistency of results, the ATS/ACCP/AASM working group recommends that clinicians consider validating the portable monitor they want to use in the settings in which they work. If portable monitoring is to be part of a clinician's practice of sleep medicine, it is recommended that a clinician ask the following three questions when choosing a particular type of monitor:

1. What is the purpose of the monitor? The ATS/ACCP/AASM working group review shows that there are more high-quality studies demonstrating an ability of portable monitoring either to exclude or to confirm a diagnosis of sleep apnea than to do both, especially at a single RDI level.
2. In what setting should the monitor be used? Portable monitors would be used ideally in the unattended (home) setting. The theoretical advantages are decreased costs and the ability to study more patients in a timely fashion. However, the majority of studies having a high level of evidence and high quality rating on portable monitors for which a sensitivity and specificity were reported have been performed in the attended setting (level I or II evidence and quality rating of a or b, 23 of 36 attended studies vs 4 of 13 unattended studies), and data proving cost effectiveness of portable monitoring are lacking. Therefore, the utility of portable monitors in an unattended setting has not been thoroughly established yet. There are, however, potential advantages to the use of portable monitors in an attended setting, primarily because of the decreased technician time needed for setup and scoring, which could prompt a clinician to use these devices instead of full polysomnography.
3. What type of monitor should be used? Available devices range from ones that perform full polysomnography to limited monitoring with flow, effort, and oximetry to simple oximetry. Full polysomnography has the potential advantage of measuring sleep parameters. Limited monitoring is simpler to perform but has the advantage of the actual detection of flow and oximetry, similar to full polysomnography. Oximetry has the advantage of simplicity. This review found large numbers of high-quality studies for both type 3 and 4 monitors.

Overall, the most consistent, high-quality data were for type 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.

5.0. Directions/Need for Future Research

This review has highlighted several areas that are necessary to address in future studies (sections 1.1.3, 1.3, 1.4, 1.5, 2.3, 2.4, 3.4, and 4.2.4) and that have been discussed under general comments and limitations in section 4.3.

5.1. Populations of Patients Not Yet Studied

To date, almost all studies on portable monitoring have been conducted in sleep clinic/laboratory populations with a high pretest probability, who are composed predominately of white men who have little or no comorbidity. Thus, it is not possible to

generalize findings from the studies that were reviewed to other groups or types of patients. Future studies should address the current gap in knowledge and should include more diverse populations of patients, as follows:

1. Primary care populations;
2. Subjects with important comorbidities;
   - COPD
   - heart failure
   - stroke
   - severe hypertension
3. Ethnic populations other than white;
4. Women.

In addition, future studies should address the utility of using clinical prediction algorithms in combination with portable monitoring, which to date has only been done in one study.[35]

5.2. Recommendations for Study Methods

Four key areas should be addressed by investigators comparing portable monitors with a reference standard such as polysomnography (Table 11). Investigators should document any perceived or actual bias that could result if the funding for the study originated with industry or if a manufacturer of a portable monitoring device paid them a consulting fee. Recruitment should be of consecutive subjects from a pool that is not subject to selection bias by the investigators. If common comorbidities are included, subjects should be defined clearly and stratified into separate groups. The sample size should be sufficient

Table 11—*Recommended Methods for Studies on Portable Monitoring for Diagnosing Sleep Apnea*

Design Feature	Key Features	Important Features
Study—general	1. Funding source 2. Investigator's possible conflict of interest	
Study population	1. Consecutive sample of subjects (prospective), not subject to selection bias by investigators 2. Broad spectrum of subjects: Wide range of pretest probability Adequate representation of both sexes Ethnicity of subjects described Clearly defined co-morbidities to allow for separate analysis	1. Description of the study setting 2. Eligibility criteria clearly stated 3. Number of subjects, clearly described, that: Were eligible to participate Refused to participate Dropped out after the study started Completed the study protocol Had incomplete or missing data
Polysomnography	1. Interpreters blind to results of subjects' portable monitoring and clinical information 2. Decision to perform polysomnography should not be influenced by results of portable monitoring 3. Preselect the criteria for a positive result (avoid *post hoc* analysis)	1. Follow recommended measurement and scoring methods[1] 2. Clear description of how breathing events were defined and scored 3. Oximeter sampling rate should be sufficient and clearly stated 4. Random assignment of subjects to polysomnography or portable monitoring for the first test (in nonsimultaneous studies) 5. Clear report of the prevalence of sleep apnea for whatever threshold(s) used to define it 6. Clear description of the equipment used and manufacturer
Portable monitoring	1. Interpreters blind to results of subjects' polysomnography and clinical information 2. Decision to perform portable monitoring should not be influenced by results of polysomnography 3. Preselect the criteria for a positive result (avoid *post hoc* analysis)	1. Evaluation of the portable monitor simultaneous with polysomnography and in an unattended setting 2. Clear description of how breathing events were defined and scored. 3. If manual scoring or editing was used, inter- and intra-scorer reliability should be reported 4. Night-to-night repeatability reported as mean differences and limits of agreement[29] 5. 95% confidence intervals reported for estimates of sensitivity, specificity and likelihood ratios 6. Details about the percentage of patients who meet the criteria for a positive or negative result, the percentage of nondiagnostic results, and the percentage of false results 7. Clear description of the equipment (including software version) used and manufacturer 8. Failure rate or percent of studies that had to be repeated because of poor quality/technical failure

to make the results representative of the population, and a power analysis based on a clearly stated hypothesis should be performed prior to the study. The study population should be described well enough to allow readers to determine whether the study subjects are similar enough to their own patient population to justify using the results in clinical practice. Details about the performance and scoring of the polysomnogram and the portable monitor should be sufficient to allow a reader to replicate the study and to ensure that all important sources of bias were controlled for. Table 11 lists the key and important features that should be expected by associate editors, reviewers, and readers of clinical studies on portable monitoring devices for sleep apnea.

Currently, polysomnography scoring varies from center to center. The AASM recommendations for research may be used as a guide.[1] The oximeter sampling rate should be sufficient to record the oxygen saturation signal accurately and should always be stated. One of the recommended means of measuring oronasal airflow should be used. It was well-recognized in the studies reviewed that the oronasal measurement was the one most subject to data loss. The definition of hypopnea should also follow the AASM recommendations for polysomnography.[1] Although an esophageal catheter is preferred to measure respiratory effort-related arousals, this is generally uncomfortable and impractical for portable studies. Respiratory effort-related arousals are included in the research definition of AHI, and this type of respiratory arousal also pertains to the upper airway resistance syndrome.[5,6] The relationship of this standardized research approach to outcomes is presently unknown, as is the clinically relevant AHI threshold target for any selected outcome.

The scoring of the portable monitor also should be standardized and clearly stated. The relationship between the portable monitor and the polysomnogram should be established with a simultaneous study in the sleep laboratory. This will provide the basis by which to judge the ideal result that could be achieved in an unattended setting. The portable monitor RDI and polysomnographic AHI thresholds should be selected prior to the study as the primary outcome variable on which to base sensitivity and specificity. The AASM recommended criterion of an AHI of ≥ 5 is appropriate for a clinical setting.[1] If an exploratory study is to be performed to determine the RDI threshold, a second study should be performed to test the accuracy of that threshold. Studies should be blinded to the scorer and should be detailed in the written methodology.

5.3. Recommendations for Study Design

5.3.1. Comparison of portable monitoring to polysomnography

It is recommended that the items in the above table be addressed if using a study design in which portable monitoring results are compared with polysomnography findings. In addition, the study design should address how to adjust unattended portable monitoring results that are compared to a separate night of polysomnography for the night-to-night variability that is observed with repeated sleep studies.[84,85] There are known and unknown variables that likely contribute to night-to-night variability[84] but are often difficult to identify, such as body position and sleep stage, particularly REM sleep. Although there is no absolute way to adjust for these variables, one approach would be to establish the night-to-night variability of the polysomnogram and of the attended polysomnogram to unattended portable monitoring in a given patient group. This could be accomplished by conducting studies on at least three separate nights, two using polysomnography and one using an unattended portable monitor not simultaneous with a polysomnogram. The time intervals between studies should be similar. The variability between polysomnogram nights 1 and 2 would be compared to the variability between the unattended portable monitoring night and each of the polysomnogram nights in the same patients. Ideally, a second unattended portable monitoring study also should be conducted to determine the night-to-night variability of the unattended portable monitor. If the variability is similar between the various comparisons of data gathered on the nights of unattended portable monitoring and those conducting an attended polysomnogram, then it is likely that differences between the portable monitor and the polysomnogram are the result of night-to-night variability. If there is a discrepancy, the magnitude of the discrepancy represents the loss of sensitivity and/or specificity from using the portable monitor. This approach also would provide information on data loss with an unattended portable monitor compared to a polysomnogram.

The order in which patients complete polysomnography and portable monitoring studies should be randomly assigned to avoid a possible order effect (*eg*, the first-night effect that is seen in patients undergoing polysomnography). In addition, it is strongly recommended that all studies use a position monitor to report the total, supine, and nonsupine AHIs and RDIs. This would help to determine how much of the night-to-night variability was the result of differences in sleeping position.

5.3.2. Alternative study designs

Apart from the technical and design factors discussed, the utility of any approach depends on the ultimate outcome of the diagnosis. If patient outcomes when using the portable monitor are similar to those when using polysomnography, the trade-off is potentially justifiable even if the sensitivity and specificity are reduced. Determining what outcomes to measure is a challenge however. Symptoms such as sleepiness are difficult or cumbersome to measure objectively. Subjective measures (*eg*, with sleepiness questionnaires or quality-of-life questionnaires) are open to a spurious result if the subject realizes that an intervention is either a placebo or a sham. For this reason, subjects should be debriefed as to whether they considered themselves to be in a placebo or intervention group. Outcomes such as reductions in traffic accidents, strokes, and myocardial infarctions may take years and large study populations to assess. Health status and health-care resource utilization are cumbersome measures, and, while they are applicable to groups of subjects, they are difficult to use in individual patients.

Another useful approach will be to assess diagnosis and management strategies with decision branches that include both polysomnography and portable monitoring. Each strategy could be investigated to determine whether outcomes such as symptoms, health status, health-care utilization, or cost-effectiveness are comparable for each branch of a given algorithm and could be considered a validation of a clinical practice guideline. For example, study limbs might include the following: (1) split-night studies; (2) attended portable monitoring followed by unattended autotitration to determine the effective CPAP pressure; (3) split-night attended portable monitoring; or (4) patients with high pretest probability who are not undergoing any diagnostic study but are proceeding directly to CPAP titration either in the laboratory or unattended with an autotitration device. The number of possible approaches and decision branches may be extensive, and subjects who fail to improve at any point would be offered alternative evaluations. In this approach, portable monitoring is not assessed as a stand-alone test but as a component of a broader strategy. Those branches that lead to comparable outcomes could be considered equal if they were powered sufficiently to establish equivalence.

There is a need to establish research priorities. The clinical demand for tests to investigate possible sleep-disordered breathing is rising.[86] As is apparent from the findings of this report, the most urgent need seems to be for additional high-quality studies to clarify the performance of portable monitors in the unattended setting. The salient point related to portable monitors is the ability to understand their utility in the setting of intended use. Successful decision making regarding the appropriate targeted outcomes combined with methods of maximum diagnostic and management efficiency seem to be the most crucial for the future research course of portable monitoring for patients with sleep-disordered breathing.

References

1 American Academy of Sleep Medicine. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research; the report of an American Academy of Sleep Medicine task force. Sleep 1999; 22:667–689

2 Young T, Palta M, Dempsey J, et al. The occurrence of sleep-disordered breathing among middle-aged adults. N Engl J Med 1993; 328:1230–1235

3 Young T, Evans L, Finn L, et al. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. Sleep 1997; 20:705–706

4 Ferber RA, Millman RP, Coppola MP, et al. ASDA standards of practice: portable recording in the assessment of obstructive sleep apnea. Sleep 1994; 17:378–392

5 American Sleep Disorders Association. Practice parameters for the indications for polysomnography and related procedures: Polysomnography Task Force, American Sleep Disorders Association Standards of Practice Committee. Sleep 1997; 20:406–422

6 Chesson AL, Ferber RA, Fry JM, et al. The indications for polysomnography and related procedures. Sleep 1997; 20: 423–487

7 Ross SD, Allen IE, Harrison KJ, et al. Systematic review of the literature regarding the diagnosis of sleep apnea: evidence report/technology assessment No. 1. Rockville, MD: Agency for Health Care Policy and Research; February 1999; AHCPR Publication No. 99–002

8 Ross SD, Sheinhait IA, Harrison KJ, et al. Systematic review and meta-analysis of the literature regarding the diagnosis of sleep apnea. Sleep 2000; 23:519–532

9 Sackett DL, Strauss SE, Richardson WS, et al. Evidence-based medicine: how to practice and teach EBM. 2nd ed. Edinburgh, Scotland, UK: Churchill Livingstone, 2000

10 Rees K, Wraith PK, Berthon-Jones M, et al. Detection of apnoeas, hypopnoeas and arousals by the AutoSet in the sleep apnoea/hypopnoea syndrome. Eur Respir J 1998; 12:764–769

11 Gugger M. Comparison of ResMed AutoSet (version 3.03) with polysomnography in the diagnosis of the sleep apnoea/hypopnoea syndrome. Eur Respir J 1997; 10:587–591

12 Mayer P, Meurice J-C, Philip-Joet F, et al. Simultaneous laboratory-based comparison of ResMed Autoset™ with polysomnography in the diagnosis of sleep apnoea/hypopnoea syndrome. Eur Respir J 1998; 12:770–775

13 Kiely J, Delahunty C, Matthews S, 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; 9:2360–2364

14 Bradley P, Mortimore I, Douglas N. Comparison of polysomnography with ResCare Autoset in the diagnosis of the sleep apnoea/hypopnoea syndrome. Thorax 1995; 50:1201–1203

15 Bagnato M, Nery L, Moura S, et al. Comparison of AutoSet and polysomnography for the detection of apnea-hypopnea events. Braz J Med Biol Res 2000; 33:515–519

16 Gugger M, Mathis J, Bassetti C. Accuracy of an intelligent

CPAP machine with in-built diagnostic abilities in detecting apneas: a comparison with polysomnography. Thorax 1995; 50:1199–1201
17 Fleury B, Rakotonanahary D, Hausser-Hauw C, et al. A laboratory validation study of the diagnostic mode of the Autoset™ system for sleep-related respiratory disorders. Sleep 1996; 19:502–505
18 Redline S, Tosteson T, Boucher MA, et al. Measurement of sleep-related breathing disturbances in epidemiologic studies: assessment of the validity and reproducibility of a portable monitoring device. Chest 1991; 100:1281–1286
19 Fry J, DiPhillipo MA, Curran K, et al. Full polysomnography in the home. Sleep 1998; 21:635–642
20 Portier F, Portmann A, Czernichow P, et al. Evaluation of home versus laboratory polysomnography in the diagnosis of sleep apnea syndrome. Am J Respir Crit Care Med 2000; 162:814–818
21 Zucconi M, Ferini-Strambi L, Castronovo V, et al. An unattended device for sleep-related breathing disorders: validation study in suspected obstructive sleep apnoea syndrome. Eur Respir J 1996; 9:1251–1256
22 Emsellem H, Corson W, Rappaport B, et al. Verification of sleep apnea using a portable sleep apnea screening device. South Med J 1990; 83:748–752
23 Man G, Kang B. Validation of a portable sleep apnea monitoring device. Chest 1995; 108:388–393
24 Ancoli-Israel S, Mason W, Coy T, et al. Evaluation of sleep disordered breathing with unattended recording: the Nightwatch System. J Med Eng Technol 1997; 21:10–14
25 Mykytyn IJ, Sajkov D, Neill AM, et al. Portable computerized polysomnography in attended and unattended settings. Chest 1999; 115:114–122
26 Parra O, Garcia-Esclasans N, Montserrat JM, et al. Should patients with sleep apnoea/hypopnoea syndrome be diagnosed and managed on the basis of home sleep studies? Eur Respir J 1997; 10:1720–1724
27 Ballester E, Solans M, Vila X, et al. Evaluation of a portable respirator recording device for detecting apneas and hypopnoeas in subjects from a general population. Eur Respir J 2000; 16:123–127
28 Ficker JH, Wiest GH, Wilpert J, et al. Evaluation of a portable recording device (Somnocheck®) for use in patients with suspected obstructive sleep apnoea. Respiration 2001; 68:307–312
29 White D, Gibb T, Wall J, et al. Assessment of accuracy and analysis time of a novel device to monitor sleep and breathing in the home. Sleep 1995; 18:115–126
30 Whittle AT, Finch SP, Mortimore IL, et al. Use of home sleep studies for diagnosis of the sleep apnoea/hypopnoea syndrome. Thorax 1997; 52:1068–1073
31 Verse T, Pirsig W, Junge-Hulsing B, et al. Validation of the POLY-MESAM seven-channel ambulatory recording unit. Chest 2000; 117:1613–1618
32 Orr WC, Eiken T, Pegram V, et al. A laboratory validation study of a portable system for remote recording of sleep-related respiratory disorders. Chest 1994; 105:160–162
33 Baltzan M, Verschelden P, Al-Jahdali H, et al. Accuracy of oximetry with thermistor (OxiFlow) for diagnosis of obstructive sleep apnea and hypopnea. Sleep 2000; 23:61–69
34 Stoohs R, Guilleminault C. MESAM 4: an ambulatory device for the detection of patients at risk for obstructive sleep apnea syndrome (OSAS). Chest 1992; 101:1221–1227
35 Gurubhagavatula I, Maislin G, Pack AI. An algorithm to stratify sleep apnea risk in a sleep disorders clinic population. Am J Respir Crit Care Med 2001; 164:1904–1909
36 Vazquez J, Tsai W, Flemons W, et al. Automated analysis of digital oximetry in the diagnosis of obstructive sleep apnoea. Thorax 2000; 55:302–307
37 Yamashiro Y, Kryger MH. Nocturnal oximetry: is it a screening tool for sleep disorders? Sleep 1995; 18:167–171
38 Wiltshire N, Kendrick A, Catterall J. Home oximetry studies for diagnosis of sleep apnea/hypopnea syndrome: limitation of memory storage capabilities. Chest 2001; 120:384–389
39 Schafer H, Ewig S, Hasper E, et al. Predictive diagnostic value of clinical assessment and nonlaboratory monitoring system recordings in patients with symptoms suggestive of obstructive sleep apnea syndrome. Respiration 1997; 64:194–199
40 Gyulay S, Olson LG, Hensley MJ, et al. A comparison of clinical assessment and home oximetry in the diagnosis of obstructive sleep apnea. Am Rev Respir Dis 1993; 147:50–53
41 Golpe R, Jimenez A, Carpizo R, et al. Utility of home oximetry as a screening test for patients with moderate to severe symptoms of obstructive sleep apnea. Sleep 1999; 22:932–937
42 Williams AJ, Yu G, Santiago S, et al. Screening for sleep apnea using pulse oximetry and a clinical score. Chest 1991; 100:631–635
43 Esnaola S, Duran J, Infante-Rivard C, et al. Diagnostic accuracy of a portable recording device (MESAM IV) in suspected obstructive sleep apnea. Eur Respir J 1996; 9:2597–2605
44 Koziej M, Cieslicki J, Gorzelak K, et al. Hand-scoring of MESAM 4 recordings is more accurate than automatic analysis in screening for obstructive sleep apnoea. Eur Respir J 1994; 7:1771–1775
45 Chiner E, Signes-Costa J, Arriero J, et al. Nocturnal oximetry for the diagnosis of the sleep apnoea syndrome: a method to reduce the number of polysomnographies? Thorax 1999; 54:968–971
46 Bonsignore G, Marrone O, Macaluso C, et al. Validation of oximetry as a screening test for obstructive sleep apnea syndrome. Eur Respir J 1990; 11(suppl):542S–544S
47 Rauscher H, Popp W, Zwick H. Quantification of sleep disordered breathing by computerized analysis of oximetry, heart rate and snoring. Eur Respir J 1991; 4:655–659
48 Cooper B, Veale D, Griffiths C, et al. Value of nocturnal oxygen saturation as a screening test for sleep apnea. Thorax 1991; 46:586–588
49 Levy P, Pepin J, Deschaux-Blanc C, et al. Accuracy of oximetry for detection of respiratory disturbances in sleep apnea syndrome. Chest 1996; 109:395–399
50 Pepin J, Levy P, Lepaulle B, et al. Does oximetry contribute to the detection apneic events?: mathematical processing of the Sa_{O_2} signal. Chest 1991; 99:1151–1157
51 Olson LG, Ambrogetti A, Gyulay SG. Prediction of sleep-disordered breathing by unattended overnight oximetry. J Sleep Res 1999; 8:51–55
52 Zamarron C, Romero P, Rodriguez J, et al. Oximetry spectral analysis in the diagnosis of obstructive sleep apnoea. Clin Sci 1999; 97:467–473
53 Series F, Marc I, Cormier Y, et al. Utility of nocturnal home oximetry for case finding in patients with suspected sleep apnea hypopnea syndrome. Ann Intern Med 1993; 119:449–453
54 Douglas N, Thomas S, Jan M. Clinical value of polysomnography. Lancet 1992; 339:347–350
55 Rauscher H, Popp W, Zwick H. Model for investigating snorers with suspected sleep apnea. Thorax 1993; 48:275–279
56 Ryan PJ, Hilton MF, Boldy DAR, et al. Validation of British Thoracic Society guidelines for the diagnosis of sleep apnoea/hypopnea syndrome: can polysomnography be avoided? Thorax 1995; 50:972–975

57 Issa F, Morrison D, Hadjuk E, et al. Digital monitoring of sleep-disordered breathing using snoring sound and arterial oxygen saturation. Am Rev Respir Dis 1993; 148:1023–1029
58 Zamarron C, Romero PV, Gude F, et al. Screening of obstructive sleep apnoea: heart rate spectral analysis of nocturnal pulse oximetric recording. Respir Med 2001; 95: 759–765
59 Reda M, Gibson GJ, Wilson JA. Pharyngoesophageal pressure monitoring in sleep apnea syndrome. Otolaryngol Head Neck Surg 2001; 125:324–331
60 Claman D, Murr A, Trotter K. Clinical validation of the Bedbugg™ in detection of obstructive sleep apnea. Otolaryngol Head Neck Surg 2001; 125:227–230
61 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1:307–310
62 Streiner DL, Norman GR. Health measurement scales: a practical guide to their development. 2nd ed. New York, NY: Oxford University Press, 1995
63 Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282:1061–1066
64 Jaeschke R, Guyatt G, Sackett DL. Users' guide to the medical literature: how to use an article about a diagnostic test; A. Are the results of the study valid? JAMA 1994; 271:389–391
65 Jaeschke R, Guyatt G, Sackett DL. Users' guide to the medical literature: how to use an article about a diagnostic test; B. What are the results and will they help me care for my patients? JAMA 1994; 271:703–707
66 West S, King V, Carey TS, et al. Systems to rate the strength of scientific evidence: evidence report/technology assessment No. 47. Rockville, MD: Agency for Healthcare Research and Quality; April 2002; AHRQ Publication No. 02-E016
67 Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol 1991; 44:763–770
68 Acebo C, Watson R, Bakos L, et al. Sleep and apnea in the elderly: reliability and validity of 24-hour readings in the home. Sleep 1991; 14:56–64
69 Ancoli-Israel S, Kripke D, Klauber M, et al. Sleep-disordered breathing in community-dwelling elderly. Sleep 1991; 6:486–495
70 Farney R, Walker L, Jensen R, et al. Ear oximetry to detect apnea and differentiate rapid eye movement (REM) and non-REM (NREM) sleep: screening for the sleep apnea syndrome. Chest 1986; 89:533–539
71 Gyulay S, Gould D, Sawyer B, et al. Evaluation of a microprocessor-based portable home monitoring system to measure breathing during sleep. Sleep 1987; 10:130–142
72 Hida W, Shindoh C, Miki H, et al. Prevalence of sleep apnea among Japanese industrial workers determined by a portable sleep monitoring system. Respiration 1993; 60:332–337
73 Lojander J, Salmi T, Maasilta P. Reproducibility of oximetry with a static charge-sensitive bed in evaluation of obstructive sleep apnoea. Clin Physiol 1998; 18:225–233
74 Polo O, Brissaud I, Sales B, et al. The validity of the static charge sensitive bed in detecting obstructive sleep apnea. Eur Respir J 1988; 1:330–336
75 Roche F, Gaspoz J, Court-Fortune I, et al. Screening of obstructive sleep apnea syndrome by heart rate variability analysis. Circulation 1999; 100:1411–1415
76 Rollheim J, Osnes T, Miljeteig H. The sites of obstruction in OSA, identified by continuous measurements of airway pressure and flow during sleep: ambulatory versus in-hospital recordings. Clin Otolaryngol 1999; 24:502–506
77 Salmi T, Partinen M, Hyyppa M, et al. Automatic analysis of static charge sensitive bed (SCSB) recordings in the evaluation of sleep-related apneas. Acta Neurol Scand 1986; 74: 360–364
78 Stoohs R, Guilleminault C. Investigations of an automatic screening device (MESAM) for obstructive sleep apnoea. Eur Respir J 1990; 3:823–829
79 Svanborg E, Larsson H, Carlsson-Nordlander B, et al. A limited diagnostic investigation for obstructive sleep apnea syndrome: oximetry and static charge sensitive bed. Chest 1990; 98:1341–1345
80 Tvinnereim M, Mateika S, Cole P, et al. Diagnosis of obstructive sleep apnea using a portable transducer catheter. Am J Respir Crit Care Med 1995; 152:775–779
81 Chediak AD, Acevedo-Crespo JC, Seiden DJ, et al. Nightly variability in the indices of sleep-disordered breathing in men being evaluated for impotence with consecutive night polysomnograms. Sleep 1996; 19:589–592
82 Tashkandi YN, Rowley JA, Badr MS. Effect of gender on PSG indices of sleep-disordered breathing [abstract]. Am J Respir Crit Care Med 2000; 161:A680
83 Redline S, Tishler PV, Hans MG, et al. Racial differences in sleep-disordered breathing in African-Americans and Caucasians. Am J Respir Crit Care Med 1997; 155:186–192
84 Le Bon O, Hoffmann G, Tecco J, et al. Mild to moderate sleep respiratory events: one negative night may not be enough. Chest 2000; 118:353–359
85 Littner MR. The diagnosis of OSA: where do we draw the line [editorial]? Chest 2000; 118:286–288
86 Namen AM, Dunagan DP, Fleischer A, et al. Increased physician-reported sleep apnea: the National Ambulatory Medical Care Survey. Chest 2002; 121:1741–1747