Diagnostic accuracy of a portable recording device (MESAM IV) in suspected obstructive sleep apnoea

S. Esnaola*, J. Durán**, C. Infante-Rivard+, R. Rubio**, A. Fernández+

Diagnostic accuracy of a portable recording device (MESAM IV) in suspective sleep apnoea. S. Esnaola, J. Durán, C. Infante-Rivard, R. Rubio, A. Fernández. ©ERS Journals Ltd 1996.

ABSTRACT: This study evaluated the accuracy of a portable recording device (MESAM IV) in identifying obstructive sleep apnoea (OSA).

The MESAM IV system measures arterial oxygen saturation (S_{a,O_2}), heart rate, snoring sounds and body position, and allows both automatic and manual scoring of the recordings. Nocturnal polysomnography and MESAM IV recordings were performed simultaneously in 150 patients with suspected OSA, and were analysed blindly by a different observer. Patients with an apnoea-hypopnoea index (AHI) greater than or equal to 10 were diagnosed as having OSA. In the evaluation of the discriminatory ability of MESAM IV scores, the cut-off point was set to minimize first the exclusion of truly diseased patients (*i.e.* false-negative interpretations), and then the confirmation of nondiseased subjects (*i.e.* false-positives).

When used as an exclusion test, the portable device reached a sensitivity of 0.98 and a specificity of 0.78; as a confirmation test, these values were 0.69 and 0.97, respectively. These results were achieved with manual scoring, which was superior to automatic scoring. Manual scoring was also better than automatic scoring when OSA was defined according to other threshold values (≥5, 15 and 20) for the AHI.

The combination of MESAM IV manual scores could reduce the need for diagnostic polysomnography in three quarters of the patients clinically suspected of having obstructive sleep apnoea, substantially reducing costs associated with diagnostic procedures.

Eur Respir J., 1996, 9, 2597–2605.

*Research Unit, Dept of Health, Basque Government, Spain. **Sleep Unit, Hospital Txagorritxu, Vitoria-Gasteiz, Spain. +Dept of Occupational Health, Faculty of Medicine, McGill University, Montréal, Canada.

Correspondence: S. Esnaola
Unidad de Investigación
Departamento de Sanidad
C/Duque de Wellington 2
01010 Vitoria-Gasteiz
Spain

Obstructive sleep apnoea (OSA) affects 2–4% of males and 1–2% of females in middle age [1, 2]. The condition is characterized by frequent episodes of apnoea and hypopnoea during sleep, which lead to the fragmentation of sleep and to decreases in oxyhaemoglobin saturation [3, 4]. As a consequence, patients are at increased risk of cardiovascular morbidity and mortality [5], and are more frequently involved in traffic accidents [6]. Despite the availability of therapies with proven effectiveness, such as continuous positive airway pressure [7], it is likely that OSA is underdiagnosed because the final diagnosis can only be carried out in a sleep laboratory, using nocturnal polysomnography by highly-qualified personnel [8, 9]. This has led to the development of portable recording devices, which can be used outside the sleep laboratory in an unattended setting. In spite of the growing clinical use of portable devices in the assessment of sleep apnoea, the evidence available concerning their diagnostic accuracy is inconclusive [10], and more research in this area has been suggested [11, 12].

The MESAM IV is a portable recording system which records oxygen saturation, heart rate, snoring sounds and body position. Four previous studies have evaluated the diagnostic validity of MESAM IV, with conflicting results [13–16]. The aim of the present study was to evaluate the diagnostic accuracy of the MESAM IV system in patients with clinical suspicion of OSA.

Material and methods

Patients and design

From November, 1991 to September, 1993, 152 consecutive patients with clinically suspected OSA were included in the study. Data corresponding to the first 51 patients were presented in a previous paper [16]. From the pneumology out-patient department, patients were referred to the sleep unit of the Txagorritxu Hospital, a university-affiliated general hospital in Vitoria-Gasteiz (Spain). They complained of daytime hypersomnolence, loud snoring, or were diagnosed as having polycythaemia, respiratory failure or cor pulmonale, and none of these diagnoses were explained by other causes. Data on sleep symptoms were obtained by using a standardized questionnaire. The study was approved by the hospital's Research and Ethics Committee. Conventional nocturnal polysomnography and MESAM IV recordings were performed simultaneously on all the patients. A different

observer carried out the readings of each of the two methods; the observers were not informed of the patient's health condition and were blind to the results of the other method. It was necessary to repeat two polysomnographs because the total sleep time was less than 180 min. As the MESAM IV printouts could not be interpreted in two patients due to recording problems, a total of 150 patients were evaluated.

Polysomnography

The polysomnography consisted of continuous polygraphic recording for a whole night (Duplex TR XVI; Alvar Electronic) of the electroencephalogram (EEG) (C3/A2), tibial and submental electromyograms (EMG), and electro-oculogram (EOG). Electrocardiogram (ECG) (modified V2 lead) was recorded with standard electrode placement, thoracoabdominal movement by plethysmography (Volucap; Alvar Electronics), oronasal flow by nasal and oral thermistors (Alvar Electronics), and oxygen saturation with an ear sensor (Critikon Oxyshuttle; Johnson & Johnson). Polysomnography records were scored for sleep, breathing, oxygenation, and movement in 30 s periods. Sleep data were staged (Stages I–IV and rapid eye movement (REM) sleep) according to the system of RECHTSCHAFFEN and KALES [17]. Both time in bed and recording time for polysomnography were equal for all patients, from 0:00 to 07:00 h. Polysomnographic recordings were analysed for the total sleep time. Mean sleep time for the 150 patients included was 318 (SD 63) min. An abnormal breathing event during objectively measured sleep was defined according to standard clinical criteria as either a complete cessation of airflow lasting 10 s or more (apnoea), or a discernible reduction in respiratory airflow accompanied by a decrease of 4% or more in oxyhaemoglobin saturation and/ or an arousal (hypopnoea). Apnoeas and hypopnoeas were subdivided into central, obstructive and mixed events. The apnoea-hypopnoea index (AHI) was calculated as the average number of episodes of apnoea and hypopnoea per hour of sleep.

MESAM IV

The MESAM IV is a portable device with four-channel digital recording, which has been described previously [18]. The system records heart rate, snoring, oxygen saturation and body position. The beginning of the recording is programmable with a personal computer (IBM compatible). Preparation of the computer and attaching the device in relation to the patient takes about 15 min. Recording can be scored automatically, by means of software included in the system, or manually, by visual evaluation of printouts. In the present study, version 3.21 of the system software was used.

Automatic analysis of MESAM IV calculates three indices. The heart rate variation index (HRVI) expresses the number of sudden increases in heart rate from the baseline per hour of analysis time. The oxygen desaturation index (ODI) expresses the number of oxygen desaturations equal to or greater than 4% per hour of analysis. The desaturation phase continues until saturation has again risen by 90% of the difference between the base level and the desaturation level. The

intermittent snoring index (ISI) is a measure of intermittent snoring sound generated by the test subject. These indices are not integrated in a final score but are presented separately.

Manual analysis aimed to summarize the changes in the heart rate, oxygen saturation and breathing sounds in a single index, which would allow the identification of the occurrence of apnoea or hypopnoea with greater accuracy than when using one of the components alone. The manual scoring analysis was made by visual evaluation of printouts at 10 min intervals ("full disclosure") (fig. 1). After a review of the whole night's recording in 2 h periods, a minimum continuous episode of 5 h was selected, which included, if present, the most abnormal section. In practice, it was found that by selecting the period from 01:00 to 06:00 h, the above criterion was achieved in 95% of the patients. The same period was used both for manual and automatic analysis. Two manual scores (MS) were developed using criteria based on those proposed by PENZEL [18]. The manual scores were calculated by a two-step process. In the first stage, the recordings of each of the three variables (*i.e.* heart rate, oxygen saturation and snoring sounds) were analysed separately. For each one, the observer marked the changes related to the possible occurrence of respiratory events, using the following criteria: for heart rate, an increase from an immediate previous baseline of 10% or more; for oxygen saturation, a drop in the level of arterial oxygen saturation (Sa,O_2) from baseline equal or greater than 4% and lasting longer than 10 s; for snoring sounds, the occurrence of at least three snores separated by intervals of 10–120 s with an absence or attenuation of snoring sounds (fig. 1). Both for Sa,O_2 and heart recordings, evident changes according to the above cited criteria were evaluated with the naked eye, and a ruler was used to measure the magnitude of doubtful changes. In the second stage, the results from the previous stage were used to calculate two manual score indices. In the two-channel manual score index (MS2), an event was defined as the simultaneous occurrence of changes in at least two of the three variables. In the case of the three-channel manual score index (MS3), an event was defined as the simultaneous occurrence of changes in all three variables. The MS2 and MS3 indices were calculated as the number of the respective events per hour of analysis time. Manual analysis, after printing of the total recording, took the technician an average of 30 min per subject.

Analysis

To measure overall agreement between the AHI and each of the scores of the MESAM IV, we calculated the intraclass correlation coefficient, which indicates how much of the total variation of the measurements can be attributed to differences between the subjects rather than to differences between the methods, and is equal to 1 when the two methods achieve exactly the same values [19]. The nature and extent of disagreement between the AHI and each of the MESAM IV scores was evaluated using the average difference ($\bar{d}$) between the measurements, and the limits of agreement ($\bar{d} \pm 1.96\ s_d$) where s_d is the standard deviation of the differences [20]. The analysis of the diagnostic accuracy of the
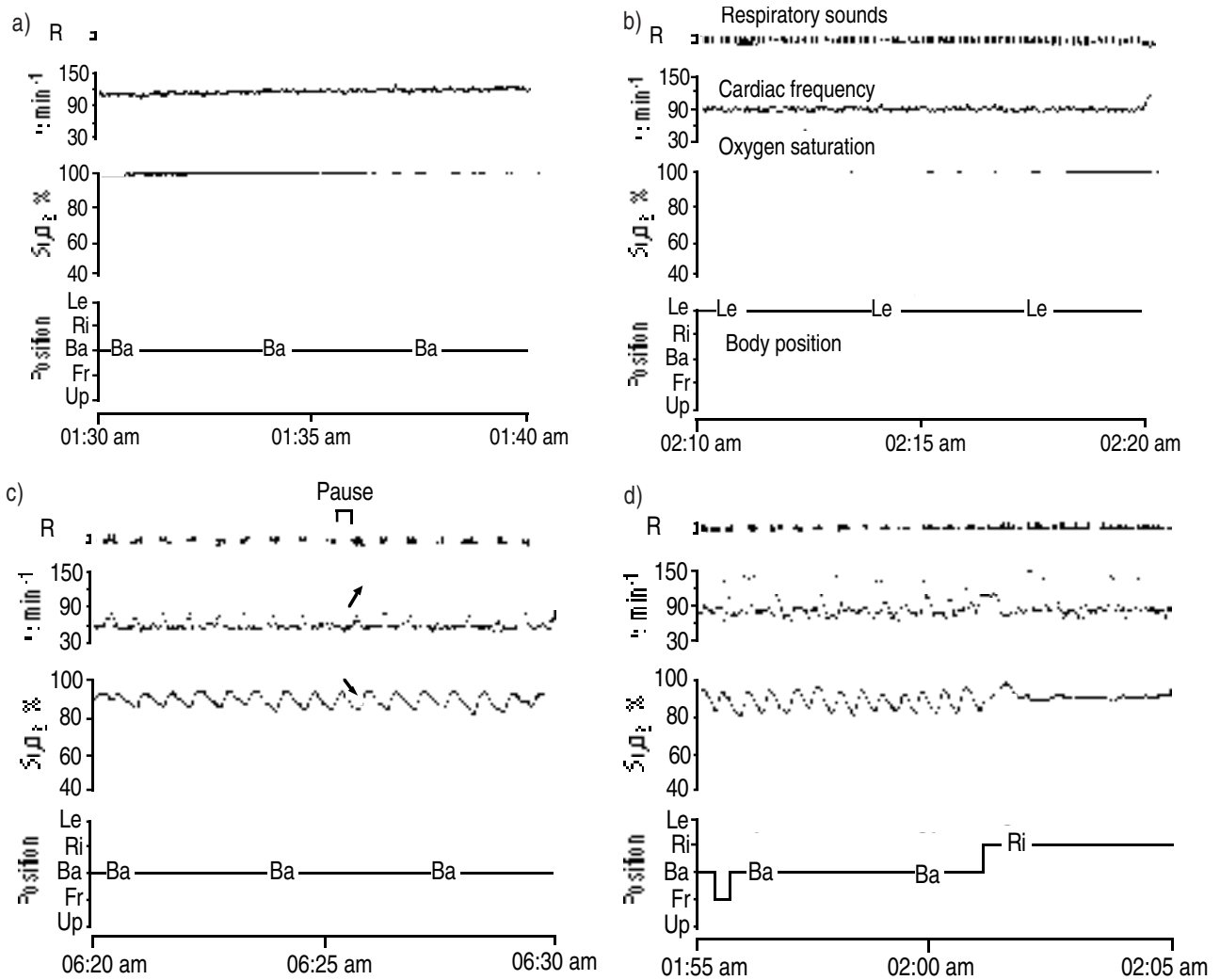
Fig. 1. – Graphic recording of the MESAM IV for visual analysis. Four examples of MESAM IV recording are presented as 10 min periods. Trace represents (from the top to the bottom) respiratory sounds (R), cardiac frequency (f_c), oxygen saturation (S_{a,O_2}), and body position. The body position channel indicates whether the subject is lying on his back (Ba), left side (Le), right side (Ri), frontal position (Fr), or is in an upright position (Up). a) Normal recording. No evidence of snoring or significant changes in heart rate or S_{a,O_2}. b) Normal subject but a heavy snorer. The number of scores can be detected on the graph. The respiratory rate can be estimated (about 12 per minute). c) Subject with OSA. Seventeen "apnoeas" can be detected from the breaks between snores. The same changes can be seen in the heart rate and the S_{a,O_2}. It is thus possible to calculate both the two channel and the three channel manual score indices. d) The same person as in graph c), where we can see how he stops having "apnoeas" and continues snoring.

MESAM IV scores was made for each of the definitions of OSA corresponding to conventional cut-off values for the AHI (≥ 5, 10, 15 and 20). The results of the present paper will focus mainly on those obtained using a threshold value of 10 for the AHI.

The discriminatory ability of the MESAM IV scores for the diagnosis of OSA was evaluated using receiver operating characteristic (ROC) curves [21]. The ROC curve shows the trade-off between sensitivity (true-positive proportion) and 1-specificity (false-positive proportion). The larger the area under the ROC curve, the more valid the diagnostic test in comparison with the gold standard (in this case conventional nocturnal polysomnography). Positive and negative predictive values were also estimated. The validity indices obtained from each of the ROC curves were estimated at cut-off points selected using two criteria. The first was the use of the MESAM IV as an exclusion test, *i.e.* to rule out the presence of OSA, sensitivity was set high at 0.97: although

most cases with OSA would be recognized as such by the MESAM IV, a relatively large proportion (in comparison with a rule-in threshold) would be false-positives. The second criterion was based on the use of the MESAM IV as a confirmation test, *i.e.* to rule-in the presence of OSA: specificity was set high at 0.97 yielding few false-positives, but relatively more false-negatives than with the rule-out criterion.

Because we do not know the probability distribution of the cut-off points obtained, the standard error of the selected cut-off points is not readily available. Furthermore, the choice of the optimal cut-off point based on the values of the study patients tends to overestimate the validity indices, the bias being important when small samples are involved. To address these two issues, the bootstrap method [22, 23] was used to estimate the standard error of the optimal cut-off point, and the bias and standard error of the corresponding validity indices (for more details see the Appendix).

Results

Patients had an average age of 57 (SD 11) yrs and 134 (89%) were males. Forty nine patients (33%) had at least a moderate chronic obstructive pulmonary disease (COPD) (FEV1 <70% of predicted value [24]). According to the results of polysomnography, 90 patients (60%) had an AHI ≥10, and were diagnosed as having OSA (table 1).

Table 1. – Characteristics of patients

	AHI ≥10 (n=90)		AHI <10 (n=60)	
Male gender n (%)	81	(90)	53	(88)
Moderate-severe COPD n (%)	30	(33)	19	(32)
Right ventricular hypertrophy (by ECG) n (%)	37	(41)	14	(25)
Breathing pauses(%)[†‡]	62	(70)	21	(35)
Moderate-severe diurnal hypersomnia(%)[†§]	48	(53)	7	(12)
Age yrs	58±11		57±10	
Body mass index kg·m^{-2}	31±5		28±4	
Diastolic blood pressure mmHg	84±11		80±12	
Oxygen desaturation index (ODI)	46±25		22±19	
Heart rate variation index (HRVI)	26±24		9±12	
Intermittent snoring index (ISI)	39±23		23±13	
Two-channel manual score index (MS2)	41±19		10±11	
Three-channel manual score index (MS3)	23±19		1.0±2.5	
Apnoea-hypopnoea index (AHI)	43±24		2.1±2.2	

Data are presented as absolute values, and percentage in parenthesis, or mean±SD. [†]: evaluated by questionnaire; [‡]: witnessed respiratory pauses reported by the room-mate or by other person living with the patient; [§]: tendency to fall asleep during daytime, excluding the period after lunch ("siesta"), almost every day or every day. COPD: chronic obstructive pulmonary disease; ECG: electrocardiography.

Table 2. – Agreement between AHI and MESAM IV scores

	Correlation coefficient			
	Intraclass	Product-moment	Bias*	Limits of agreement[†]
Automatic scores				
ODI	0.47 (0.33–0.58)	0.51 (0.38–0.62)	-94	-61 41.7
HRVI	0.36 (0.22–0.49)	0.43 (0.28–0.55)	74	-46 60.6
ISI	0.44 (0.30–0.56)	0.48 (0.34–0.59)	-59	-56 43.8
Manual scores				
MS2	0.72 (0.63–0.79)	0.73 (0.65–0.80)	-1.8	-38 34.8
MS3	0.57 (0.63–0.79)	0.75 (0.66–0.81)	12.6	-24 48.7

Values in parentheses are 95% confidence intervals. *: average difference ($\bar{d}$)=(AHI) - (MESAM IV score). [†]: $\bar{d}$±1.96 s_d, where s_d is the standard deviation of the differences. ODI: oxygen desaturation index; HRVI: heart rate variation index; ISI: intermittent snoring index; MS2: two-channel manual score index; MS3: three-channel manual score index; AHI: apnoea-hypopnoea index.
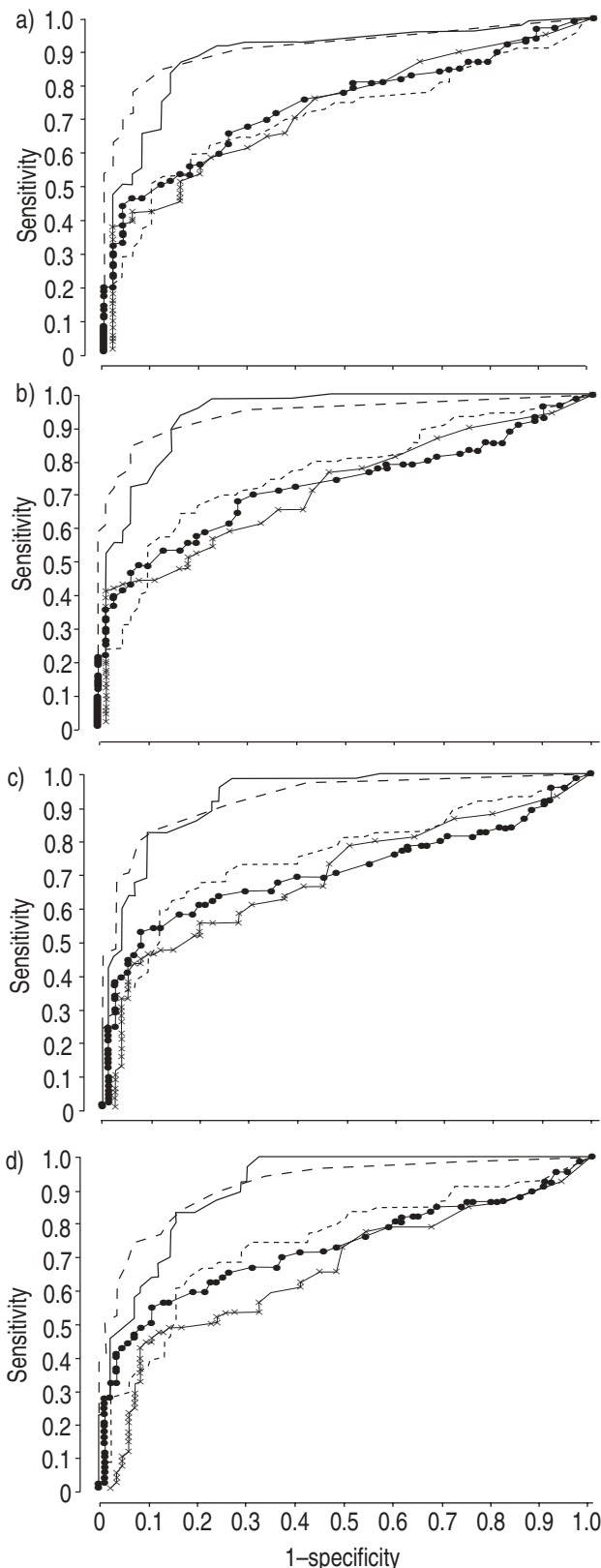

Fig. 2. – Receiver operating characteristic (ROC) curves for MESAM IV scores in obstructive sleep apnoea defined according to different threshold values for the apnoea-hypopnea index (AHI). a) AHI ≥5; b) AHI ≥10; c) AHI ≥15; d) AHI ≥20. ——: two channel manual score index (MS2); – – –: three channel manual score index (MS3); - - - -: oxygen desaturation index (ODI); –×–×–: heart rate variation index; –●–●–: intermittent snoring index (ISI).

Table 3. – Area under the receiver operating characteristic (ROC) curve (and 95% confidence interval in parenthesis) of MESAM IV scores in obstructive sleep apnoea (OSA)

	Diagnostic criterion for OSA			
	AHI ≥5	AHI ≥10	AHI ≥15	AHI ≥20
Automatic scores				
ODI	0.710	0.764	0.760	0.748
	(0.627–0.793)	(0.688–0.841)	(0.681–0.839)	(0.666–0.830)
HRVI	0.730	0.718	0.700	0.665
	(0.649–0.811)	(0.637–0.799)	(0.615–0.785)	(0.574–0.755)
ISI	0.744	0.726	0.712	0.728
	(0.666–0.821)	(0.647–0.806)	(0.625–0.798)	(0.641–0.814)
Manual scores				
MS2	0.893	0.939	0.932	0.910
	(0.837–0.949)	(0.899–0.979)	(0.893–0.970)	(0.865–0.954)
MS3	0.914	0.943	0.931	0.918
	(0.869–0.960)	(0.907–0.979)	(0.891–0.971)	(0.873–0.962)

For definitions see legend to table 2.

Table 2 summarizes the findings on the agreement between the AHI and the MESAM IV scores. The agreement was clearly greater in the MS2 than in the automatic scoring. The width of the limits of agreement indicates great variability in the differences between the AHI and the MESAM IV scores.

Figure 2 presents the ROC curves for the MESAM IV scores, using different diagnostic criteria for OSA. As evidenced by the values for the area under the ROC curve (table 3), irrespective of the diagnostic criteria, the discriminatory ability of the manual scores was greater than that observed with automatic scores.

Tables 4–7 show the validity indices corresponding to the optimal cut-off points selected from each ROC curve using two decision criteria. When considering the use of the MESAM IV as an exclusion test (rule-out

criterion), the best results were obtained with the MS2; for instance, when considering an AHI ≥10 as diagnostic of OSA, only two cases out of the 90 diagnosed by polysomnography were missed by the MESAM IV (sensitivity 0.98; 95% confidence interval (95% CI) 0.95–1.00), while 47 out of the 60 patients without OSA were correctly identified (specificity 0.78; 95% CI 0.68–0.89). Thus, OSA was correctly ruled-out in 47 patients of the 49 with negative MESAM IV results (negative predictive value 0.96; 95% CI 0.90–1.00). When using the MESAM IV as a confirmation test (rule-in criterion) the MS3 obtained the best validity indices; using the same diagnostic criterion for OSA (AHI ≥10), only one subject among the 60 without OSA was wrongly diagnosed as having OSA by the MESAM IV (specificity 0.98; 95% CI 0.95–1.00), while 62 of the 90 affected patients were recognized as such (sensitivity 0.69; 95%

Table 4. – Validity indices of MESAM IV scores in obstructive sleep apnoea (AHI ≥5)

	Cut-off point	Sens.	Spec.	PPV	NPV
Confirmation test*					
Automatic scores					
ODI	60	0.29	0.96	0.94	0.41
HRVI	28	0.37	0.96	0.95	0.44
ISI	41	0.44	0.96	0.96	0.47
Manual scores					
MS2	39	0.51	0.96	0.96	0.50
MS3	10	0.66	0.96	0.87	0.80
Exclusion test**					
Automatic scores					
ODI	2	0.95	0.04	0.66	0.29
HRVI	1	0.95	0.10	0.67	0.50
ISI	2	0.97	0.08	0.67	0.57
Manual scores					
MS2	4	0.97	0.18	0.70	0.75
MS3	0	1.00	0.00	0.66	-

*: selected cut-off point corresponding to the specificity closest to 0.97; **: selected cut-off point corresponding to the sensitivity closest to 0.97. PPV: positive predictive value; NPV: negative predictive value. Sens.: sensitivity; Spec.: specificity. For further definitions see legend to table 2.

Table 5. – Validity indices of MESAM IV scores in obstructive sleep apnoea (AHI ≥10)

	Cut-off point	Sens.	Spec.	PPV	NPV
Confirmation test*					
Automatic scores					
ODI	66	0.23	0.98	0.95	0.46
HRVI	27	0.42	0.97	0.95	0.53
ISI	45	0.40	0.97	0.95	0.52
Manual scores					
MS2	39	0.56	0.97	0.96	0.59
MS3	11	0.69	0.98	0.98	0.67
Exclusion test**					
Automatic scores					
ODI	2	0.97	0.07	0.61	0.57
HRVI	1	0.94	0.08	0.61	0.50
ISI	3	0.97	0.10	0.62	0.67
Manual scores					
MS2	14	0.98	0.78	0.87	0.96
MS3	1	0.96	0.69	0.83	0.91

*: selected cut-off point corresponding to the specificity closest to 0.97; **: selected cut-off point corresponding to the sensitivity closest to 0.97. For definitions see legends to tables 2 and 4.

CI 0.59–0.78). Thus, OSA was correctly ruled-in in 62 of the 63 patients with positive MESAM IV results (positive predictive value 0.98; 95% CI 0.95–1.00). By combining MS2 and MS3, figure 3 shows that 38 patients (25%) of the 150 studied were not appropriately classified by the MESAM IV: among those, 26 had OSA and 12 did not.

The bootstrap procedure was used to study the stability of the selected cut-off points (see Appendix). When considering an AHI ≥10 as diagnostic of OSA, and using

Table 6. – Validity indices of MESAM IV scores in obstructive sleep apnoea (AHI ≥15)

	Cut-off point	Sens.	Spec.	PPV	NPV
Confirmation test*					
Automatic scores					
ODI	61	0.36	0.96	0.90	0.60
HRVI	63	0.12	0.97	0.82	0.53
ISI	50	0.39	0.97	0.94	0.61
Manual scores					
MS2	50	0.47	0.97	0.95	0.65
MS3	14	0.68	0.97	0.96	0.75
Exclusion test**					
Automatic scores					
ODI	2	0.96	0.05	0.50	0.57
HRVI	1	0.93	0.07	0.50	0.50
ISI	3	0.96	0.08	0.51	0.67
Manual scores					
MS2	17	0.96	0.76	0.80	0.95
MS3	1	0.97	0.58	0.70	0.96

*: selected cut-off point corresponding to the specificity closest to 0.97; **: selected cut-off point corresponding to the sensitivity closest to 0.97. For definitions see legends to tables 2 and 4.

Table 7. – Validity indices of MESAM IV scores in obstructive sleep apnoea (AHI ≥20)

	Cut-off point	Sens.	Spec.	PPV	NPV
Confirmation test*					
Automatic scores					
ODI	66	0.28	0.96	0.86	0.63
HRVI	76	0.01	0.98	0.33	0.55
ISI	55	0.33	0.98	0.92	0.64
Manual scores					
MS2	51	0.46	0.98	0.94	0.69
MS3	26	0.52	0.98	0.95	0.71
Exclusion test**					
Automatic scores					
ODI	2	0.96	0.05	0.45	0.57
HRVI	1	0.93	0.06	0.44	0.50
ISI	2	0.96	0.05	0.45	0.57
Manual scores					
MS2	17	0.97	0.70	0.72	0.97
MS3	1	0.97	0.52	0.63	0.96

*: selected cut-off point corresponding to the specificity closest to 0.97; **: selected cut-off point corresponding to the sensitivity closest to 0.97. For definitions see legends to tables 2 and 4.

Fig. 3. – Classification of patients according to the MESAM IV manual scores. Data are expressed as absolute value, and percentage in parenthesis. MS2: two-channel manual score index; MS3: three-channel manual score index; AHI: apnoea-hypopnea index.

the rule-out criterion, the mean value of the selected cut-off point for MS2 was 14.2 (95% CI 13.4–15.1). According to the rule-in criterion, the mean value of the cut-off point for MS3 was 10.2 (95% CI 7.4–13.8). The estimated bias for sensitivity and specificity in both cases was less than ±0.004.

In order to assess reliability of the manual scoring method, 126 randomly selected records were blindly re-analysed by the same observer. Another set of 56 randomly selected records was reanalysed by a sleep technician. Agreement for the classification of patients according to the rule-in and the rule-out criteria was evaluated by Cohen's kappa (κ) [25]. When OSA was defined by an AHI ≥10, intraobserver agreement values were 0.96 (95% CI 0.91–1.00) for the MS2 and 0.984 (95% CI 0.95–1.00) for the MS3. The values of κ for the interobserver agreement were 0.76 (95% CI 0.56–0.96) for MS2 and 0.86 (95% CI 0.72–0.99) for MS3.

Discussion

In this study, visual evaluation of MESAM IV recordings achieved a high diagnostic accuracy in patients with clinical suspicion of OSA; results were not as good using automatic scoring. The results of the present study are similar to those of ROOS et al. [14] and KOZIEJ et al. [15]. The estimated validity for the automatic scores in the present study contrasts with the results obtained by STOOHS and GUILLEMINAULT [13] where the ODI attained a sensitivity of 0.92 and a specificity of 0.97, whereas the corresponding values for the HRVI were 0.58 and 0.32, and for the ISI 0.96 and 0.27. As regards the ODI, the results obtained by STOOHS and GUILLEMINAULT [13] are also in contrast with those found in previous studies which have evaluated the diagnostic validity of oximetry:

when an AHI ≥10 was used as a diagnostic criterion for OSA, sensitivity values of more than 0.90 were accompanied by a specificity no greater than 0.53 [12].

The version of the MESAM IV system software used does not appear to explain the differences found between the studies being considered. When the recordings obtained in the present study were reanalysed with software version 3.1, which was that used in the STOOHS and GUILLEMINAULT [13] study, the estimated validity indices showed no appreciable changes for any of the automatic scores. The criteria for the choice of recording time to be analysed were different in the studies under consideration. STOOHS and GUILLEMINAULT [13] selected the analysis period by using the lights out and lights on times recorded in the patient sleep logs, with the additional support of the visual evaluation of the heart rate recording. ROOS et al. [14] used the EEG recording to select the period corresponding to the sleep stage. However, this criterion could not be applied if the MESAM IV were used on its own. In this study, we chose a minimum standard continuous recording time of 5 h, which is usually the most representative for an overall view, and which in 95% of cases was from 01:00 to 06:00 h. As far as the type of patients is concerned, while STOOHS and GUILLEMINAULT [13] included patients with sleep/wake-related complaints, in the other studies the patients studied were clinically suspected of having OSA.

Potential limitations of the present study must be addressed. Bias can arise when the determinations of the status of each patient, according to the "gold standard" and the studied test, are not made independently of each other [26]. In this study, this bias was prevented by the simultaneous application of both recording methods, and by an independent assessment of each recording in the absence of knowledge of the characteristics of patients. The reliability of manual scoring of MESAM IV printouts could compromise their accurate use. The manual scoring procedure used in the present study was developed in a standardized fashion, and based on objective criteria. Results of inter- and intraobserver agreement suggest that manual scoring of the MESAM IV is a procedure which can be used in a way which is highly reproducible. This conclusion is supported by the observations made in the study by ROOS et al. [14], in which the three observers achieved very similar validity indices. The present study was limited to patients with suspected OSA. Thus, the results observed do not allow assessment of the diagnostic accuracy of the MESAM IV when applied to subjects with other characteristics, i.e. asymptomatic patients, or in other settings. This last caveat is of special relevance when considering the use of the MESAM IV outside the sleep laboratory in an unattended setting.

The MESAM IV system has several limitations, which should be pointed out. Given the absence of measurements for the direct determination of sleep staging, the calculated scores refer to the selected recording time, which does not always correspond to the sleep stage. In some subjects with auricular fibrillation or who suffer from arrhythmias during sleep, it is not possible to interpret the heart rate recording. Drops in Sa,O_2 of more than 50% are not registered, which makes manual analysis difficult and leads to the underestimation of events in the automatic analysis. Sensors can become disconnected, producing artefacts which are not always detected by the software. Furthermore, with the MESAM IV, apnoeas are not identified; their occurrence is deduced from the changes in Sa,O_2, heart rate and respiratory sound. However, these measurements can be altered by causes other than OSA, such as COPD, artefacts and REM sleep. Moreover, snoring in subjects who snore a great deal, without clear interruptions, gives rise to difficulties in interpretation.

Standard polysomnography continues to be the accepted diagnostic method used for obstructive sleep apnoea [8–12]. Nevertheless, growing clinical awareness of the frequency and potential consequences associated with OSA have led to the search for methods that allow its diagnosis at a lower cost. For this purpose, portable recording devices are being widely used in the diagnostic evaluation of OSA outside the sleep laboratory, in an unattended mode. The American Sleep Disorders Association has recently reviewed the publications on unattended recording devices [11]. The MESAM IV system was included among the level III devices, that allow for the assessment of cardiorespiratory variables but do not allow determination of wakefulness and sleep stages. Most of these level III devices have been evaluated in an attended setting, with the implicit assumption that a good diagnostic performance in this setting is a necessary condition for it to be so outside the laboratory, where there is usually no sleep technician. It was from this perspective that our study was proposed. Its results indicate that manual scoring of the MESAM IV applied in the sleep laboratory cannot substitute for polysomnography, but it can play a complementary diagnostic role in patients who are clinically suspected of having OSA. The MESAM IV could be used as the first examination to be carried out in such patients. Were OSA defined by an AHI ≥10, those patients classified as negative according to the MS2 would not require a subsequent polysomnographic study. The selected cut-off point could vary in each setting, in accordance with the expected prevalence of OSA and the balance required between sensitivity and specificity. Patients classified as positive according to the MS3 could be considered as suffering from OSA. In those patients for whom it was decided to start treatment with continuous positive airway pressure, the polysomnography needed to adjust the pressure level would serve to confirm the diagnosis, so that no patient mistakenly classified as positive would receive the treatment. By making use of this strategy and in accordance with the results of the present study, manual scoring with MESAM IV could eliminate the need for diagnostic polysomnography in three quarters of the patients clinically suspected of having OSA.

The complexity and costs of the alternative diagnostic strategies must be considered. On the one hand, the introduction of manual analysis of MESAM IV requires specific training. In our experience, the expertise needed to carry out the manual analysis by a sleep technician can be acquired after a training period of 30 h. Furthermore, given that the current version of the MESAM IV system does not allow the monitoring of the recording during the study time, the frequency of lost data when used in a sleep unit could be higher than that observed

in this study. The experience of an ongoing epidemiological study of OSA [27], in which the MESAM IV system is being used at home in an unattended mode, indicates that the proportion of evaluations that should be repeated would be less than 10%. On the other hand, the "partially attended" nature of the MESAM IV recording allows for improved comfort during sleep and reduces the likelihood of the "first night" effects. Moreover, the attachment of the device requires considerably less time and produces less discomfort than polysomnography. As far as monetary costs are concerned, we have compared the proposed diagnostic strategy (first MESAM IV system for all patients, and thereafter polysomnography for those with doubtful MESAM IV results) with the conventional approach of submitting all patients to polysomnography. Assuming a cost of US \$546 for each polysomnographic study, and of US\$155 for each MESAM IV; if 10% of additional MESAM IV evaluations need to be repeated, and polysomnography is required in 25% of patients (classified as "doubtful"), the cost of the proposed strategy ($155+ 0.1 \times 155 + 0.25 \times 546 = 307$ per patient) would produce a net cost reduction of 44% (546 - 307/546).

In summary, the manual scoring of a MESAM IV recording as applied in the sleep laboratory is a valuable tool in the diagnostic assessment of patients with suspected obstructive sleep apnoea. Manual scoring of MESAM IV is clearly better than automatic scoring in terms of agreement with apnoea-hypopnoea index, and to discriminate patients with obstructive sleep apnoea. As indicated by its low intra- and interobserver variability, manual scoring analysis of MESAM IV is a reliable procedure which can be used to identify those patients who should undergo polysomnography. Future studies should assess the diagnostic accuracy of the MESAM IV system in an unattended setting and in asymptomatic subjects.

Appendix

The bootstrap is a computer-intensive method which, as applied here, involved drawing two sets of 1,000 random samples of 150 patients each (with replacement) from the original sample of 150 patients. From the first set of bootstrap samples, an optimal cut-off point (mean and standard error) and its corresponding validity indices were estimated. By application of the optimal cut-off point of set one to the second set of bootstrap samples, an estimate of the validity indices was obtained. In this way, the procedure imitates the development of a diagnostic rule from a set of patients and the subsequent assessment of performance in another group. Bias was estimated as the average difference between the two estimates of the validity indices.

References

1. Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S. The occurrence of sleep-disordered breathing among middle-aged adults. *N Engl J Med* 1993; 328(17): 1230–1235.
2. Partinen M, Telakivi T. Epidemiology of obstructive sleep apnea syndrome. *Sleep* 1992; 15: S1–S4.
3. Guilleminault C, van den Hoed J, Mitler M. Clinical overview of the sleep apnea syndromes. *In*: Guilleminault C, Dement W, eds. Sleep Apnea Syndromes. New York, Alan R. Liss, 1978; pp. 1–11.
4. Fletcher EC, Costarango C, Miller T. The rate of fall of arterial oxyhemoglobin saturation in obstructive sleep apnea. *Chest* 1988; 96: 717–722.
5. Partinen M, Guilleminault C. Daytime sleepiness and vascular morbidity at seven year follow-up in obstructive sleep apnea patients. *Chest* 1990; 97: 27–32.
6. Findley LJ, Levinson MP, Bonnie RJ. Driving performance and automobile accidents in patients with sleep apnea. *Clin Chest Med* 1992; 13(3): 427–435.
7. Engleman HM, Martin SE, Deary IJ, Douglas NJ. Effect of continuous positive airway pressure treatment on daytime function in sleep apnea/hypopnea syndrome. *Lancet* 1994; 343: 572–575.
8. Martin RJ, Block AJ, Cohn MA, *et al*. Indications and standards for cardiopulmonary sleep studies. *Sleep* 1985; 8: 371–379.
9. American Thoracic Society. Indications and standards for cardiopulmonary sleep studies. *Am Rev Respir Dis* 1989; 139: 559–568.
10. Stiller RA, Strollo PJ, Sanders MH. Unattended recording in the diagnosis and treatment of sleep-disordered breathing: unproven accuracy, untested assumptions, and unready for routine use. *Chest* 1994; 105: 1306–1309.
11. Standards of Practice Committee of the American Sleep Disorders Association. Practice parameters for the use of portable recording in the assessment of obstructive sleep apnea. *Sleep* 1994; 17: 372–377.
12. Ferber R, Millman R, Coppola M, *et al*. Portable recording in the assessment of obstructive sleep apnea. *Sleep* 1994; 17: 378–392.
13. Stoohs R, Guilleminault C. MESAM 4: an ambulatory device for the detection of patients at risk for obstructive sleep apnea syndrome (OSAS). *Chest* 1992; 101(5): 1221–1227.
14. Roos M, Althaus W, Rhiel C, Penzel T, Peter JH, von Wichert P. Comparative use of MESAM IV and polysomnography in sleep-related respiratory disorders. *Pneumologie* 1993; 47 (Suppl. 1): 112–118.
15. Koziej M, Cieslicki JK, Gorzelak K, Sliwinski P, Zielinski J. Hand-scoring of MESAM 4 recordings is more accurate than automatic analysis in screening for obstructive sleep apnoea. *Eur Respir J* 1994; 7: 1771–1775.
16. Durán Cantolla J, Esnaola Sukia S, Rubio Aramendi R, Egea Santaolalla C. Validez de un sistema de registro portátil (MESAM IV) para el diagnóstico del síndrome de apnea del sueño. *Arch Bronconeumol* 1994; 30: 331–338.
17. Rechtschaffen A, Kales AA. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Washington, DC, US Government Printing Office, NIH Publication No. 204, 1979.
18. Penzel T, Althaus W, Meinzer K, Peter JEI, Wichert P. A device for ambulatory heart rate, oxygen saturation and snoring recording. 13th Annual International Conference IEEE EMBS. *New Frontiers of Biomedical Enginering* 1991; 13: 1616–1617.
19. Kelsey J, Thompson WD, Evans AS. Methods in

observational epidemiology. New York, NY, Oxford University Press, 1986; pp. 288–293.

20. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–310.

21. Swets JA, Picket RM. Evaluation of diagnostic systems. New York, NY, Academic Press, 1982.

22. Efron B, Gong G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *Am Stat* 1983; 37: 36–48.

23. Linnet K, Brandt E. Assessing diagnostic tests once an optimal cut-off point has been selected. *Clin Chem* 1986; 32: 1341–1346.

24. American Thoracic Society. Lung function testing: selection of reference values and interpretative strategies. *Am Rev Respir Dis* 1991; 144: 1202–1218.

25. Armstrong BK, White E, Saracci R. Principles of exposure measurement in epidemiology. Oxford, Oxford University Press, 1994; pp. 104–109.

26. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficiency of diagnostic tests. *N Engl J Med* 1978; 299: 926–930.

27. Esnaola S, Durán J, Rubio R, Iztueta A. Prevalence of obstructive sleep apnoea in the male population of Vitoria-Gasteiz (Spain) (Abstract). *Eur Respir J* 1995; 8 (Suppl. 19): 436s.