Accuracy of Oximetry with Thermistor (OxiFlow) for Diagnosis of Obstructive Sleep Apnea and Hypopnea[1,2]

Marc A. Baltzan MD,[3] Paul Verschelden MD, Hamdan Al-Jahdali MD, Allen E. Olha MSc, R. John Kimoff MD[4]

[1]From the Respiratory Division, McGill University Health Centre (Royal Victoria Hospital Site), Montreal, Quebec, Canada, *[2]This study was supported by Nellcor Puritan Bennett,* *[3]Dr. Baltzan is supported by the Medical Research Council of Canada,* *[4]Dr. Kimoff is a clinical research scholar of the Fonds de la recherche en santé du Québec*

Objectives: To evaluate the diagnostic accuracy for obstructive sleep apnea and hypopnea (OSAH) of the OxiFlow (OF) device which combines oximetry with recording of thermistor airflow.
Design & Setting: Patients scheduled for overnight diagnostic polysomnography (PSG) were studied with OF either simultaneously during laboratory PSG (L-OF, n=86), at home on a separate night (H-OF, n=66), or both (n=55).
Patients: 97 patients with suspected OSAH, of whom 40 had OSAH defined as an apnea-hypopnea index (AHI) of more than 15 events per hour of sleep on PSG.
Interventions: NA
Measurements & Results: The automated respiratory disturbance index (RDI) generated by the OF software considerably underestimated the AHI by PSG for both L-OF and H-OF. Altering the parameters for hypopnea identification by the software did not improve this. Visual inspection of the computerized OF tracings added considerable diagnostic information, but a manual count of RDI during visual review overestimated AHI. For the identification of cases vs. non-cases of OSAH, receiver operating characteristic area-under-the-curve statistics ranged from 0.77 – 0.90 for L-OF and from 0.71 – 0.77 for H-OF. Combining automated analysis with subsequent visual inspection of OF tracings yielded an overall sensitivity of 86% and specificity of 74% for the diagnosis of OSAH during H-OF recordings. Analysis of potential technician time saved indicated a benefit from the use of OF.
Conclusions: OF has diagnostic utility for the identification of OSAH. However, because of hardware and software limitations, it is unclear whether this device is superior to oximetry alone.
Key words: Sleep apnea; polysomnography; portable monitoring; ambulatory diagnosis; oximetry; cost-effectiveness; diagnostic accuracy

INTRODUCTION

OBSTRUCTIVE SLEEP APNEA AND HYPOPNEA (OSAH) OCCURS IN 3% TO 24% OF THE NORTH AMERICAN POPULATION[1,2] and is the most common reason for referral for diagnostic polysomnography.[3-12] There is a need for ready access to diagnostic testing for OSAH, which in turn requires low-cost and accurate diagnostic methods if wide-spread diagnosis and treatment of this prevalent disorder are to be feasible.[13-15] The accuracy of oximetry alone is variable depending on the population studied and the criteria used for diagnosis[3-8,16] and is not formally recommended as a sole criterion for diagnosis.[16] More complex home recording equipment that uses at least four data acquisition channels may provide high diagnostic specificity.[16] Yet, this requires greater involvement of a technician for set up of the patient, has a rate of data loss of 4% to 24% when used in the home, and consequently has an increased cost per study in comparison to oximetry. Further evaluation of current and new technology is needed.[14,16,17] A home diagnostic device that would not require a technician for patient set-up may expedite the diagnosis of OSAH. The use of oximetry combined with thermistor airflow monitoring (OxiFlow) could represent a next step in improving the clinical utility of oximetry with only a small increase in complexity and cost.

The aim of this study was therefore to evaluate the diagnostic accuracy and utility of the OxiFlow (OF). The specific aims were to: compare the accuracy of OF in the sleep lab (L-OF) to simultaneous polysomnography (PSG); to compare the automated default scoring algorithm in the OF software with four alternate hypopnea scoring criteria and a manual count for estimating the apnea-hypopnea index (AHI) by PSG; to evaluate if the diagnostic device provided information superior to pulse oximetry alone based upon

Accepted for publication July 1999
Address for Correspondence: Dr. R. John Kimoff, Rm. L4.08, Royal Victoria Hospital - McGill University Health Center, 687 Pine Avenue West, Montreal, Quebec, Canada. H3A 1A1. Tel: (514) 843-1568; Fax: (514) 843-1695; E-mail: john.Kimoff@muhc.mcgill.ca

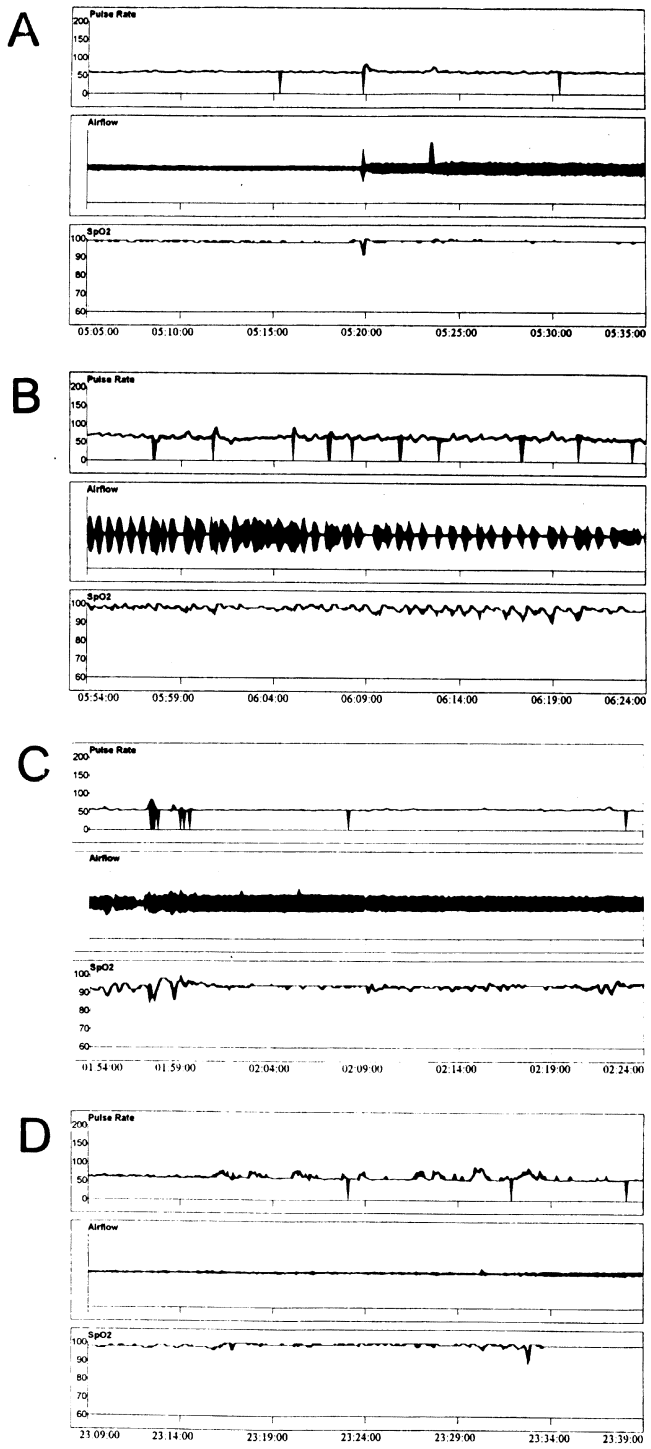

Figure 1—Representative 30 min. printouts of computer-generated OF tracings. Panel A: Subject without OSAH correctly identified as such by OF (true negative) showing non-oscillatory variability of SaO_2 and thermistor signals; Panel B: True positive subject demonstrating typical oscillations in SaO_2 and airflow; Panel C: False negative subject showing minimal fluctuation in thermistor signal, and minor changes in SaO_2; Panel D: false positive subject demonstrating a thin thermistor signal envelope which was interpreted as showing numerous events by the automated analysis software.

blinded manual review of the tracings; and to evaluate the feasibility and diagnostic utility of OxiFlow in the home set up by the patient (H-OF) compared with PSG.

METHODS

Device

The home recording device uses two sensors: a finger pulse oximeter sampling at 10 Hz, and an oronasal thermistor mounted on the upper lip. The portable memory will store up to 13 hours of continuous recording with the data stored as maximum and minimum values over six second intervals. Following acquisition, the data are downloaded to a personal computer and are displayed on the monitor as envelopes of pulse rate, airflow, and oxygen saturation. Epochs may be tailored in the report, with a maximum resolution of 30 minutes per window (Figure 1). These may be printed for a written report. No information is provided to assess respiratory effort.

Automatic scoring is provided by the OF software with criteria chosen by the operator. This automated score produces a respiratory disturbance index (RDI) that is the sum of all apneas and hypopneas divided by total recording time. Apneas were defined in all analyses as an absence of airflow signal lasting 10 seconds or more. Various criteria for identification of hypopneas were tested. In a preliminary analysis of 22 L-OF studies we determined the accuracy of 18 sets of hypopnea criteria with the flow reduction requirement ranging from 10% to 70% and the O_2 desaturation requirement ranging from 0 to 4%. It was found that varying the definition of 10 second hypopneas with decrements in airflow between 20% to 50 % and oxygen saturation between 2% to 4% provided the most stable estimates over a wide range of respiratory disturbance indices (RDI). Accordingly, results are reported for the OxiFlow software default hypopnea criteria of 50% reduction in airflow with 4% desaturation (default criteria) as well as for four alternate criteria: 25% reduction in airflow with 4% desaturation; 20% reduction in airflow with 4% desaturation; 50% reduction in airflow with 2% desaturation; and 25% reduction in airflow with 2% desaturation.

Manual Scoring

Manual scoring of the OxiFlow tracings was performed by one of the authors (RJK). These studies were scored in a blind fashion, where the scorer was presented all tracings in an order generated by a random number algorithm, blinded to all other clinical information on the study subjects. Initially, only the oximetry tracings were presented, the scorer documented a decision of positive or negative, and then the airflow tracings were revealed. The scorer was asked specifically to classify the tracing as positive for OSAH in the presence of repetitive episodes of transient

desaturation followed by a rapid return to the baseline SaO_2 level, using no minimum decrease in SaO_2 levels and no threshold, as described by Sériès et al.[3] Similar explicit rules regarding transient changes in heart rate (from the pulse oximeter) and airflow were also used. As such all repetitive variations in any of the three cardiorespiratory signals without specific threshold were counted as events. A standard data sheet required a uniform assessment of the adequacy of the tracing, and a judgment whether both airflow and oximetry were necessary in order to come to a diagnostic decision regarding the presence or absence of OSAH. Visual inspection of the tracings required a summary decision regarding case status ('OSAH' or 'no OSAH'). On recordings done simultaneously in the sleep lab, a manual count of all respiratory events was made and divided over total recording time to estimate a manual RDI. A second investigator (PV) scored a sample of the studies in parallel; interrater concordance was high with a Cohen's kappa 0.88 for the presence or absence of OSAH and a correlation of 0.81 for estimation of manual RDI.

Study Population

Consecutive patients who were scheduled to undergo nocturnal PSG in the Royal Victoria Sleep Laboratory between September, 1996 and March, 1997 with a suspicion of OSAH were approached for study. Only studies without split night protocols were included. Many cases of severe OSAH seen in the laboratory during this time period were not recruited because they were studied either with split night protocols or during the daytime; the study type was selected at the treating physician's discretion based upon the presence of severe impairment or active cardiorespiratory disease. The clinical characteristics of the subjects who enrolled were collected by interview and examination by the investigators, and included the Epworth sleepiness scale. When subjects performed a home OF recording (H-OF) they were asked to fill out a brief questionnaire asking if any events, malfunctions, or difficulties with equipment were observed; they were also asked to evaluate if they slept supine during the recording. Subjects signed informed consent, and the study was approved by the Royal Victoria Hospital Ethics Committee.

Protocol

Prior to the initiation of the protocol, pilot studies were performed to thoroughly familiarize all staff with the correct operation of the OF and interpretation of the reports. Contingent upon consent and availability of the devices, patients were recorded with OF either in the sleep lab simultaneously with PSG (L-OF), at home on a separate night (H-OF), or both. When both were obtained, H-OF and L-OF were performed within two weeks of each other in random order, unless intercurrent illness, such as upper respiratory tract infection, developed and forced a delay. All instances of use of the OxiFlow were recorded. Data on 11 subjects who did not have an interpretable overnight PSG are excluded from presentation and analysis. The reasons for these 11 exclusions were a total sleep time less than 180 minutes (n=5), the patient refused or rescheduled PSG after the H-OF (n=4), and the digital file for scoring was lost or corrupted (n=2).

For L-OF recordings the OF equipment was installed and operated by the PSG technologist. For H-OF studies, the subjects were instructed in the use of the machine on the day of the study by one of the investigators or a sleep lab technologist, during a session that lasted a maximum of 10 minutes. The subjects then installed the device themselves at home with the aid of a sheet developed for the study with several simple diagrams illustrating the procedure.

Measurements

Standard polysomnography was performed on all patients. Sleep staging was performed using standard electroencephalographic (EEG) leads (C4-A1/C3-A2); bilateral electrooculogram; chin and anterior tibialis electromyograms; airflow oro-nasal pressure cannula[18] paired with the thermistor; thoraco-abdominal movements by inductive plethysmography (Respitrace Systems, Respitrace Corp., Ardsley, NY); and arterial oxyhemoglobin saturation by finger pulse oximetry (Ohmeda Biox 3700, Ohmeda Corp., Boulder, CO). All signals were acquired on a digital data management system (Sandman, Nellcor-Puritan Bennett, Ottawa, Canada). One registered polysomnographic technologist (RPSGT) with 18 years of experience scored the studies blind of the results of OxiFlow monitoring. An apnea was defined a priori as an episode of >90% reduction in airflow lasting at least 10 seconds and a hypopnea as a 50% drop in airflow with coincident desaturation of at least 4% lasting at least 10 seconds. Obstructive sleep apnea-hypopnea was defined as an apnea-hypopnea index (AHI) of more than 15 events per hour of EEG sleep. Upper airway resistance (UAR) episodes were defined as discrete episodes of inspiratory airflow contour reduction or flattening on the nasal cannula pressure signal[18] associated less than 4% desaturation, terminating with an EEG arousal.[19] A case of increased upper airway resistance was defined as the presence of an AHI of less than 15 with repeated UAR episodes resulting in an arousal index of at least 20 events or more per hour of sleep.[19] Central apneas were scored if apneas occurred in the absence of respiratory effort or inspiratory airflow contour flattening.

Data Analysis

Demographic data are summarized with descriptive statistics with standard deviations. All OxiFlow scores in

comparison with PSG were analyzed with Pearson product-moment correlation,[20] Bland and Altman plots,[21] as well as sensitivity and specificity[22] for detecting AHI>15. Confidence intervals for proportions were computed with the mid-p exact method.[23] Standard receiver operating characteristic (ROC) analysis[24] was performed with the use of the LABROC1 program (Kurt Rossman Laboratories, University of Chicago). Statistical significance was declared at a p-value of <0.05.

Analysis of Potential Time Savings

An analysis of direct costs potentially saved in regard to technician time was performed as described previously.[25] Three strategies are compared in this group of subjects, the second and third of which are based on findings presented in the Results section. The first strategy (strategy 1) describes the time needed for a program where all patients with suspected OSAH undergo PSG. The second strategy (strategy 2) estimates the time needed for a program where all patients with suspected OSAH undergo a first OF study at home, and only those with an automated RDI between 2 and 20 by default criteria undergo subsequent PSG. The third strategy (strategy 3) estimates the time needed for cost of a program where all patients with suspected OSAH undergo a screening OF study at home, and PSG is performed only on those with both an RDI of 2 to 20 (default criteria) and where visual screening can not exclude OSAH. Disposable supply costs and equipment depreciation were not included. The calculations are presented in terms of technician time. It was assumed that each polysomnography requires four hours of technician time for the set up, supervised acquisition and scoring when a ratio of technicians to patients of one to three per night is used as has been recommended.[13] It is assumed that the same technician executes both PSG and OxiFlow studies, and on average an OxiFlow study requires a half hour of technician time per patient. Physician time to review the OF printouts requires one to two minutes per study, and was not included in the analysis. All lost OF studies were included, and for the sake of computation assumed to be diagnostic OF failures followed by PSG. Indirect costs of missed or delayed diagnosis were not calculated as these are highly variable and depend heavily upon physician judgment and the heath care system.

Table 1—Subject Characteristics

Age (yr.)	51.8 (14.6)
Female (%)	25 (26)
BMI(kg/m2)	28.4 (6.2)
Epworth Score	9.6 (5.1)
Habitual Snoring (%)	82 (84)
Witnessed Apneas (%)	57 (58)
Total sleep time (hours:minutes)	5:14 (1:11)
Sleep onset latency (minutes)	31.7 (39.4)
Stage changes	131.7 (57.4)
Awakenings	25.3 (15.2)
Arousal index	33.9 (19.2)
Percent of sleep time	
Stage 1	15.4 (12.7)
Stage 2	57.4 (14.0)
Slow wave sleep	13.8 (11.9)
REM sleep	13.4 (11.9)
Minimum SpO2	84 (7.9)
AHI	18.0 (18.9)
% above 15 events/hour	40
% above 10 events/hour	52
UARS (%)	12 (12)

Values are means (standard deviations) or numbers with percent (%). Abbreviations: AHI: apnea-hypopnea index; UARS: upper airway resistance syndrome.

RESULTS

Study Population

The population characteristics are described in Table 1. A total of 108 subjects were recruited, of which 97 had adequate PSG. Of the 97 subjects 40 (41%) met the study definition of OSAH. Twelve further cases met the criteria for increased upper airway resistance, with a mean (SD) of 45.3 (16.6) episodes of respiratory events per hour of sleep and a mean arousal index of 46.7 (13.3). No subjects had predominately central sleep apnea. Two subjects had peri-

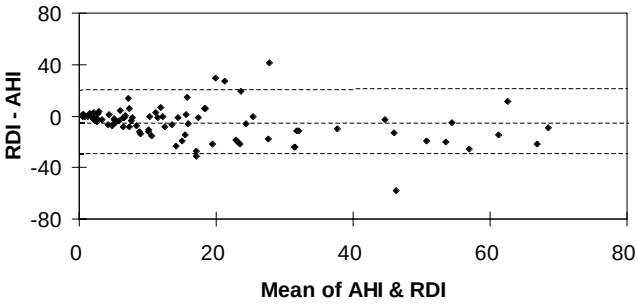

Figure 2—Bland & Altman plot for the RDI values from the default criteria automated analysis of laboratory OF studies, compared with the gold standard apnea-hypopnea index (AHI) from simultaneous polysomnography. The dashed lines represent the mean difference in measures of -5.4 (middle line) and the 2 standard deviations of the differences, 21.0 and -31.7 (lines above and below).

odic leg movements (PLM), one with and one without concomitant OSAH. Eighty-six patients were studied with OF in the sleep lab simultaneously with PSG (L-OF), 66 were studied at home on a separate night (H-OF), and 55 underwent both L-OF and H-OF.

Adequacy of Study Data

Both lab and home OF recordings provided adequate studies 92.0% and 91.3% of the time respectively. Reasons for lost data were uninterpretable thermistor probe signal (n=6), poor oximeter probe signal quality (n=3), problems in downloading the data to the analysis monitor (n=3) and the OxiFlow machine was not turned on at lights out in the lab (n=1). There was no significant difference in the number of lost or inadequate studies for L-OF vs. H-OF (odds ratio 1.12, 95% CI 0.36, 3.51, p-value 0.85). Two input cables had to be repaired during the course of the study after being damaged by patients.

Studies Performed Simultaneously with Laboratory PSG (L-OF)

The RDI provided by the automated analysis of OF data using the default criteria correlated with the AHI from simultaneous PSG with a rho=0.75. The four alternate hypopnea criteria correlation coefficients ranged from 0.71 to 0.84. Slopes and intercepts may be inferred from the Bland & Altman plot (Figure 2). All automated scoring methods tended to underestimate AHI, especially if the AHI was high (Figure 2). The manual count of all discernible events performed by one investigator during visual review of the computerized tracings correlated less well with AHI (rho=0.68) and tended to overestimate the true AHI.

Using an AHI of >15 on polysomnography as the standard to define a diagnosis of OSAH, the sensitivity and specificity of the automated OF analysis using default criteria were nevertheless 97% and 32% respectively when utilizing an automated RDI of less than 2. Specificity increased to 95% utilizing an automated RDI of greater than 20 to define a case (Table 2). However, RDI values fell below 2 or above 20 in only 43% of subjects recorded. Use of the alternate hypopnea criteria gave only marginal improvement in sensitivity and specificity. The ROC analysis area under the curve (AUC) statistic demonstrated diagnostic utility for all criteria, but no criteria were significantly superior to others, as the AUC confidence intervals demonstrated a wide overlap (Table 3, first column of AUC values).

Visual review of the OF tracings to identify patterns of repetitive O_2 desaturation, pulse rate oscillations, or variation in airflow, was felt to add useful diagnostic information in many cases. Representative tracings are shown in Figure 1. In a majority of recordings the oximetry tracings were considered sufficient for a decision regarding the presence or absence of OSAH. When oximetry alone was compared sequentially to oximetry with thermistor, the airflow signal was judged to provide useful additional information for interpretation of the record in 33% of L-OF and 31% of H-OF recordings.

Visual inspection of the tracings with a summary decision regarding case status ("OSAH" or "no OSAH") yielded a sensitivity of 100% and a specificity of 52% (Table 2). All cases meeting criteria for UARS on PSG were identi-

Table 2—Diagnostic Accuracy of OF Studies

	Lab (L-OF)		Home (H-OF)	
	Sensitivity	Specificity	Sensitivity	Specificity
Visual inspection	1.00	.52	.88	.62
Manual RDI of 20	1.00	.15	-	-
Automated RDI*				
(default criteria)				
2	.97	.32	.90	.32
10	.73	.83	.55	.88
15	.58	.93	.34	.94
20	.43	.95	.31	.97
30	.28	.98	.07	1.00

* Respiratory disturbance index used as the critical value to define 'positive'.

Table 3—Receiver Operating Characteristic Analysis

	Lab (L-OF)		Home (H-OF)	
Scoring Criteria	**AUC**	**95% CI**	**AUC**	**95% CI**
Manual Scoring	.81	.69 to .92		
Automated*				
50 and 4	.83	.75 to .91	.77	.65 to .89
25 and 4	.90	.83 to .97	.72	.60 to .84
20 and 4	.89	.83 to .95	.71	.58 to .83
50 and 2	.77	.67 to .87	.77	.65 to .88
25 and 2	.84	.76 to .92	.71	.58 to .83

*Criteria designate % decrements for hypopneas in airflow and O2 saturation, respectively.
Abbreviations: AUC: area under the curve statistic; 95% CI: 95% confidence intervals.

fied as abnormal by visual screening of the L-OF tracings. The default automated RDI for these individuals ranged from 2.9 to 12.9. If both subjects with AHI>15 and UARS were defined together as cases, this increased the specificity of visual inspection to 71%.

Comparison of PSG and L-OF Tracings

While the above analyses were performed in a blinded fashion, subsequent comparison of PSG and simultaneous L-OF tracings was performed in an attempt to identify sources of inaccuracy.

One important observation was that movement artifacts on the PSG record could appear strikingly similar on the OF tracing to true respiratory events. For example, in Fig. 1A the dip in SaO_2 in the middle of the epoch was related to body movement, but appears virtually identical to many of the "true" desaturation episodes recorded with OF. This appeared to account in part for the relatively poor specificity of visual inspection of OF tracings and the overestimation of RDI by the manual count.

Another finding was that true apneas or hypopneas could be associated with the occurrence of a pulse search artifact by the OF oximeter during a desaturation. According to the manufacturer, this would then result in the exclusion from analysis of that event. An example of this is shown in Fig. 1B, where at least eight of the ten pulse search artifacts on the upper tracing were associated with true apneas, but would have been excluded from the automated analysis. This clearly contributed to the underestimation of AHI by the automated RDI in some cases.

A further observation was that in a substantial proportion of recordings, the OF thermistor signal showed relatively little fluctuation during unequivocal hypopneas on the PSG recording (Fig. 1C). Thus the thermistor appeared relatively insensitive to changes in flow in some subjects. Due to the reliance of the OF analysis algorithm on thermistor flow, this led to the underestimation of AHI by the automated analysis. In several cases, short hypopneas appeared to be evident on the thermistor signal but were missed by the analysis algorithm. In five cases, long periods of poor quality flow signal appeared to be counted as repetitive apneas, generating a falsely high RDI (Fig. 1D).

Studies Performed on Separate Nights at Home (H-OF)

The automated RDI on a separate night at home, as scored by the default criteria, correlated less well with AHI, producing a coefficient of 0.58. The four alternate hypopnea criteria demonstrated weaker correlation, with coefficients ranging from 0.53 to 0.40. All automated scoring methods again underestimated AHI.

Sensitivity and specificity of the automated default criteria for the diagnosis of OSAH were 90 and 32% respectively for an RDI of 2, to 31 and 97% for an RDI of 20 (Table 2). For these H-OF studies, 41% of subjects demonstrated RDI values below 2 or above 20. The three false negative H-OF (AHI >15 with RDI < 2) either reported not sleeping supine at home (two subjects) or had REM-related OSAH by PSG (one subject). The alternate criteria yielded slightly worse sensitivities and specificities. The ROC plot for automated analysis is shown in Figure 3. The calculated area under the curve statistic (AUC) for all criteria again demonstrated significant diagnostic utility for OSAH, with the AUC of 0.77 for the default criteria remaining the highest for home studies (Table 3, second column of AUC values).

Visual inspection of the tracings from H-OF recordings was again felt to add useful information, and yielded a sensitivity of 88% and a specificity of 62% for a decision of positive or negative for OSAH (Table 2).

As noted, the automated OF analysis either excluded OSAH (RDI<2) or identified OSAH (RDI >20) with a high degree of accuracy in 41% of H-OF patients. We wished to assess the accuracy for all patients of a combined approach using an initial automated OF analysis followed by visual inspection of the OF tracings of the remaining 62% of subjects with 2 < RDI <20. The overall calculated sensitivity was 86% and specificity was 74% for OSAH using this combined approach on our H-OF subjects.

A night-to-night variability in the subjects' respiratory variables was observed. When both H-OF and L-OF studies were interpretable, 12 of 53 (23%) had a difference in RDI of 10 or more between the two nights, a finding that has been previously observed in patients with PSG on consecutive nights.[26]

Analysis of Potential Time Savings

Strategy 1, where all patients referred would undergo only PSG, would have required four hours of technician time per patient. With strategy 2, patients would be excluded from PSG with a default criteria RDI of either <2 (14 screened as no OSAH) or RDI >20 (14 screened as definite OSAH), while the remainder would undergo PSG. As noted above, this would have avoided PSG for 28 of 68 patients (41%, 95% confidence intervals 30 – 53%). This yielded a potential reduction of total technician time from 4.0 to 2.9 h per patient (95% confidence intervals 2.4 – 3.3 h). Strategy 3 would exclude patients as in strategy 2, with additional patients screened visually if the default RDI fell between 2 and 20 (i.e., the "combined approach" described immediately above). Inspection would have screened another 14 of these as cases of OSAH, excluding the need for PSG in 42 of 68 patients (62%, 95% confidence intervals 50 –73%). The total technician time with strategy 3 is two hours per patient (95% confidence intervals 1.6 – 2.5).

DISCUSSION

In this study we assessed the diagnostic accuracy for OSAH of OF compared with PSG during both simultaneous laboratory recording and separate home studies. We identified inaccuracies in the numerical estimation by OF of the respiratory indices determined from PSG, and our comparison of simultaneous PSG and L-OF records identified several potential explanations for these. Despite this, OF demonstrated diagnostic utility and the potential to reduce costs for case-findings in OSAH, particularly if both automated analysis and visual inspection of the data are performed.

Our study population consisted of consecutive patients with suspected OSA referred to our laboratory for overnight PSG. We believe this group is representative of the patients currently being referred to sleep disorders services for evaluation of sleep-disordered breathing (Table 1). Furthermore, the subject group was appropriate for the evaluation of a diagnostic device for OSAH in that there was a fairly even distribution of subjects with (40% OSAH, 12% UAR) and without the disorder, and there was a broad range of severity among affected individuals. The only potential concern is that very severe OSA patients may have been under-represented, in that patients with a "florid" clinical presentation of severe OSA are often evaluated in our laboratory with split night or daytime studies, which were not included in the OF study sample. However, severe cases were present in the study group, and based on the performance of OF in these subjects, it seems likely that a larger number of severe patients in the study cohort would have led to slightly improved sensitivity and specificity values for OF. Thus our figures may represent a slight underestimate in relation to the entire population of individuals referred for OSA. It should be noted that no patient in this study had predominantly central apnea and events observed by OF were assumed to be obstructive in nature. The findings cannot be extended to patients suspected of having predominantly central apnea as this was not specifically addressed in this study.

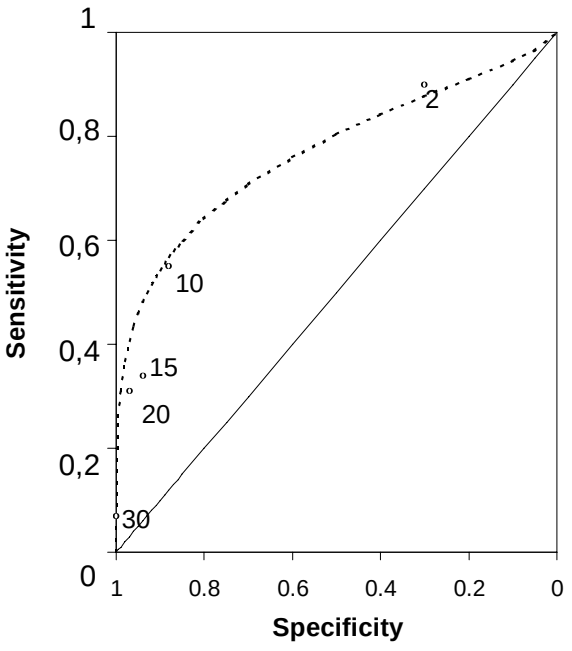

Figure 3—Receiver Operating Characteristic curve for the identification of OSAH using the automated OF analysis with default criteria from home OF recordings. The values denote RDI ranging from 2 to 30. The curve is a maximum likelihood estimate based on the data.

An important requirement for a home diagnostic device is ease of application. For home studies, our subjects received brief instruction on the use of OF and then installed themselves at home with only the aid of a diagram. The device was well accepted by most subjects, and technically inadequate or lost studies were infrequent (8.3%). Of note, these occurred with similar frequency for home and for laboratory recordings during which a trained technologist installed and operated the device. Overall, the device appears practical in terms of ease of application.

Evaluation of the numerical accuracy of the RDI generated by the OF analysis software demonstrated that the OF RDI markedly underestimated AHI from the PSG, when the same standard criteria for events were used both in the software and for scoring of the PSG. Re-analysis comparing the OF RDI with the "RDI" value calculated from the PSG (i.e., total number of events/total recording time, rather than the AHI) produced only minimal improvements in accuracy, due to a high mean sleep efficiency on PSG. We attempted to improve the accuracy of OF RDI by varying the flow and desaturation criteria for hypopneas in the analysis software. This did not produce significant improvements in accuracy.

Visual inspection of the OF tracings using an approach for identification of events similar to that described by Sériès et al.[3] was found to provide useful additional information and improved diagnostic accuracy. The fact that additional information was gained from inspection of the tracings suggests that the available data is not being optimally utilized by the software, and that modifications to the analysis algorithm might improve the automatic identification of events. However, when the manual count of respiratory events was performed during visual review of L-OF studies this was found to overestimate RDI.

The direct comparison of PSG records and simultaneous L-OF tracings identified three main factors which could account for the inaccuracies of automated and manual OF event counts. The first was the apparent poor sensitivity of the thermistor for physiologically important changes in airflow, in that the thermistor signal in some patients showed minimal fluctuation during unequivocal hypopneas on the PSG recording. In that during L-OF studies proper positioning of the thermistor was confirmed by the technologist, the insensitivity of the thermistor signal was likely

related to the recognized limitations of thermistor technology for measurement of respiratory airflow or an effect of the OF sampling and display software or both. Given that the OF automated analysis algorithm relies heavily on the thermistor for identification of events this would lead to inaccuracies in the automated analysis.

The second problematic aspect of OF software was that the loss of the pulse signal or a pulse search by the oximeter was recorded as a zero value and appeared as such on the OF tracing. Furthermore, it was not uncommon for true desaturation episodes to be associated with a pulse search at the end of the event. In that all pulse search desaturations are excluded from analysis by the OF software, many respiratory events may thus not have been identified, which would have contributed to the underestimation of AHI.

The third problematic aspect of the software was the envelope nature of the OF tracings, particularly in regard to SaO_2. The tracings consist of envelopes representing the minimum and maximum values sampled over a six-second interval for each signal. Because of this, movement artifact could appear indistinguishable from a true desaturation episode (Fig. 1). Unfortunately artifact could not be excluded either on the basis of a pulse search (cf. above), or from the thermistor signal which could show apnea-like variation with movement. The difficulty in identifying movement artifact undoubtedly contributed both to the inaccuracy of the automated RDI and to the overestimation of RDI during visual review of OF tracings.

There was some numerical discrepancy between the OF RDI values from L-OF and H-OF studies. This was likely due at least in part to night-to-night variability in sleep-disordered breathing.[26] The extent to which technical factors and issues of measurement reproducibility may have contributed is difficult to ascertain from our data. Incorrect probe placement by inexperienced patients or unobserved probe displacement could have reduced the sensitivity of H-OF recordings. Alternatively, the simultaneous use of the nasal pressure cannulae and OF thermistor during L-OF PSG studies could have influenced the RDI in that trials with patterned breathing indicated that the variation in amplitude of the thermistor airflow signal was slightly greater with the cannulae in place than without. Thus automated detection of events may have been relatively less sensitive in some H-OF recordings.

Despite the inaccuracies identified, both the automated RDI generated by OF and visual inspection of the tracings demonstrated diagnostic utility for identification of cases vs. non-cases of OSAH. Values for the ROC AUC statistic ranged from 0.77 to 0.83 for automated scoring using the default criteria automated analysis. While there is little published ROC data on OSAH diagnostic methods for comparison, AUC values in this range are observed for several commonly used radiographic diagnostic procedures.[24] A published clinical prediction index for OSAH[27] demonstrated an overall AUC value of 0.78, which is therefore comparable to our findings with OF.

Cut-off values for automated RDI could be used to yield a high sensitivity or a high specificity for both L-OF and H-OF studies (Table 2). For H-OF studies, an automated RDI <2 excluded OSAH with a false negative rate of only 4.5% (3 among 68 subjects), and an RDI >20 identified OSAH with a false positive rate of only 1.5% (1 among 68). While automated RDI values in less than half of the subjects fell either below 2 or above 20, if a combined approach was used of subsequently performing visual review of OF tracings for all subjects falling between 2 and 20, the overall calculated sensitivity of H-OF studies for OSAH was 86% with a specificity of 74%.

While these findings support the diagnostic utility of OF, does this device provide any meaningful advantage over oximetry alone? Table 4 summarizes published data on simple home recording devices. It can be seen here as well as from Table 2 that the sensitivity of all these devices can be very high, while specificity seems to peak at approximately 75%, with OF presenting no additional advantage. One potential approach to improving this could be the concurrent use of a clinical prediction rule,[27] which may increase the efficiency of a program of simplified diagnosis. A concurrent study at our institution has shown that the addition of OF information to a clinician's impression as

Table 4—'Self Set-up' Home Devices

Reference	Device	Site	n	% Lost	Sensitivity	Specificity
Douglas et al (8)	oximetry	lab	200	?	67	77
Sériès et al. (3)	oximetry	home	240	8	100	39
Yamashiro et al.(5)	oximetry	lab	300	10	90	75
Issa et al. (10)	Snoresat	lab	129	0	97	82
Present study	OxiFlow	home	66	9	86	74
		lab	86	8	98	59

rated on a visual analog scale demonstrates better diagnostic accuracy than either the OF or the impression alone (manuscript in preparation). Other diagnostic devices that employ more channels may provide stronger diagnostic accuracy, but this would be at the cost of substantial technician time for set up, scoring, and device management with potential data loss.

Our analysis of potential time savings indicates that use of OF in either of two diagnostic strategies reduces the technician time required for case-finding in OSAH. The analysis was conservative in that all failed first studies were counted as non-diagnostic and thus requiring PSG, when in reality a second OF study could be performed. Non-diagnostic PSG's also occurred, and these were not included in the analysis. Our findings indicate that the need for diagnostic PSG and labor requirement for the diagnosis of OSAH could be significantly reduced when OF is used as a diagnostic device.

In summary, our findings indicate that use of OF is a feasible method to screen patients with suspected OSAH. The RDI generated by the OF software poorly estimates AHI by PSG, which appears to be due to limitations of thermistor technology and some aspects of the display and automated analysis software. Despite this, OF generates sufficient information to discriminate between cases and non-cases with reasonable accuracy, particularly if visual inspection of the tracings is performed in tandem with automated scoring. Using this device for simplified diagnosis of patients with significant OSAH has the potential to provide technical labor cost savings. However it is not clear from our results that the use of OF is superior to oximetry alone.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the technical contributions to this study of Victoria Champagne, Nadia Brienza, France Pelligrini, Dr. John Walsh and Robert Wilson.

REFERENCES

1. Kripke DF, Ancoli-Israel S, Klauber MR, Wingard DL, Mason WJ, Mullaney DJ. Prevalence of sleep-disordered breathing in ages 40-64 years: a population-based survey. Sleep 1997;20:65-76.
2. 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;328:1230-5.
3. Series F, Marc I, Cormier Y, La Forge J. Utility of nocturnal home oximetry for case finding in patients with suspected sleep apnea hypopnea syndrome. Ann Int Med 1993;119:449-53.
4. Gyulay S, Olson LG, Hensley MJ, King MT, Allen KM, Saunders NA. A comparison of clinical assessment and home oximetry in the diagnosis of obstructive sleep apnea. Am Rev Respir Dis 1993;147:50-3.
5. Yamashiro Y, Kryger MH. Nocturnal oximetry: is it a screening tool for sleep disorders? Sleep 1995;18:167-71.
6. Levy P, Pepin JL, Deschaux-Blanc C, Paramelle B, Brambilla C. Accuracy of oximetry for detection of respiratory disturbances in sleep apnea syndrome. Chest 1996;109:395-9.
7. Ryan PJ, Hilton MF, Boldy DA, Evans A. Bradbury S, Sapiano S, Prowse K, Cayton RM. Validation of British Thoracic Society guidelines for the diagnosis of the sleep apnoea/hypopnoea syndrome: can polysomnography be avoided? Thorax 1995;50:972-5.
8. Douglas NJ, Thomas S, Jan MA. Clinical value of polysomnography. Lancet 1992;339:347-50.
9. Rauscher H, Popp W, Zwick H. Model for investigating snorers with suspected sleep apnoea. Thorax 1993;48:275-9.
10. Issa FG, Morrison D, Hadjuk E, Iyer A, Feroah T. Remmers JE. Digital monitoring of sleep-disordered breathing using snoring sound and arterial oxygen saturation. Am Rev Respir Dis 1993;148:1023-9.
11. Schafer H, Ewig S, Hasper E, Luderitz B. Predictive diagnostic value of clinical assessment and nonlaboratory monitoring system recordings in patients with symptoms suggestive of obstructive sleep apnea syndrome. Respiration 1997;64:194-9.
12. Man GC, Kang BV. Validation of a portable sleep apnea monitoring device. Chest 1995;108:388-93.
13. Whitelaw WA, Flemons WW. Polysomnography: addressing the need for standards. Can Med Assoc J 1996;155:1693-4.
14. Baumel MJ. Maislin G. Pack AI. Population and occupational screening for obstructive sleep apnea: are we there yet? Am J Respir Crit Care Med 1997;155:9-14.
15. Hida W. Kurosawa H. Miki H. Go T. Shindoh C. Kikuchi Y. Takishima T. Shirato K. [Portable home monitoring system in screening for sleep-disordered breathing]. [Japanese] Nippon Kyobu Shikkan Gakkai Zasshi - Jap J Thoracic Dis 1995;33 Suppl:46-9.
16. Ferber R, Millman R, Coppola M, Fleetham J, Murray CF, Iber C, et al. ASDA Standards of practice: portable recording in the assessment of obstructive sleep apnea. Sleep 1994;17:378-92.
17. Official Statement of the American Thoracic Society. Indications and standards for cardiopulmonary sleep studies. Am Rev Respir Dis 1989; 139:559-68.
18. Montserrat JM, Farre R, Ballester E, Felez MA, Pasto M, Navajas D. Evaluation of nasal prongs for estimating nasal flow. Am J Respir Crit Care Med 1997;155:211-5.
19. Guilleminault C. Stoohs R. Clerk A. Cetel M. Maistros P. A cause of excessive daytime sleepiness. The upper airway resistance syndrome. Chest 1993;104:781-7.
20. Armitage P, Berry G. Statistical methods in medical research. 3rd ed. Oxford: Blackwell Science, 1994.
21. Bland JM, Altman DG. Statistical methods for assessing the agreement between two methods of clinical measurement. Lancet 1986;1: 307-10.
22. Lang TA, Secic M. How to report statistics in medicine. Philadelphia: American College of Physicians, 1997:147-70.
23. Edwardes M.D. de B. The evaluation of confidence sets with application to binomial intervals. Statistica Sinica 1998;8:393-409.
24. Swets JA. Measuring the accuracy of diagnostic systems. Science 1988; 240:1285-93.
25. Drummond MF, Stoddart GL, Torrance GW. Methods for economic evaluation of health care programmes. 2nd ed. Oxford: Oxford University Press;1994.
26. Chediak AD, Acevedo-Crespo JC, Seiden DJ, Kim HH, Kiel MH. Nightly variability in the indices of sleep-disordered breathing in men being evaluated for impotence with consecutive night polysomnograms. Sleep 1996;19:589-92.
27. Maislin G, Pack AI, Dribbs NB, Simth PL, Schwartz AR, Kline LR, Schwab RJ, Dinges DF. A survey screen for prediction of apnea. Sleep 1995;18:158-66.