

# Impact of Clinical Assessment on Use of Data from Unattended Limited Monitoring as Opposed to Full-in lab PSG in Sleep Disordered Breathing

Maria Jose Masdeu, M.D.<sup>1</sup>; Indu Ayappa, Ph.D.<sup>2</sup>; Dennis Hwang, M.D.<sup>2</sup>; Anne M. Mooney, M.D.<sup>2</sup>; David M. Rapoport, M.D.<sup>2</sup>

<sup>1</sup>Pulmonary Department, Corporacio Sanitaria Parc Tauli, Universitat Autonoma de Barcelona, Sabadell, Barcelona, Spain;

<sup>2</sup>Division Pulmonary and Critical Care Medicine, New York University School of Medicine, New York, NY

SCIENTIFIC INVESTIGATIONS

**Study Objectives:** We examined agreement among multiple sleep clinicians when presented with clinical data plus the full tracings and data obtained from unattended limited monitoring (ULM) or a full polysomnography (PSG).

**Methods:** Subjects included 66 patients with complaints of sleep disordered breathing (SDB) and 19 volunteers willing to undergo 2 nights of ULM followed by PSG. Two assessment packages were created for each subject with identical clinical history (Hx) and ARES Symptom Questionnaire, plus the electronic record of signals collected on the ARES Unicorder (Hx+ULM) or on the PSG (Hx+PSG). Data were presented to 4 sleep-trained clinicians for diagnosis and treatment recommendation. For agreement on diagnosis and treatment, comparisons were made between clinicians using ULM or PSG, and within clinicians comparing both techniques.

**Results:** For diagnosis, agreement between pairs of clinicians using Hx+PSG ranged from 74% to 86% and 66% to 85% when using Hx+ULM. For treatment, agreement using Hx+PSG ranged from 74% to 86% and 58% to 77% when

using Hx+ULM. Agreement between clinicians was highest in the subjects with the highest RDI and fell off markedly at the lowest RDI, irrespective of whether the clinicians used the Hx+PSG or Hx+ULM. This pattern was also seen for the decisions made by an individual clinician using Hx+ULM vs. Hx+PSG.

**Conclusion:** Our data show that sleep clinicians have significant disagreements for diagnosis even when presented with the "gold standard" of a PSG and clinical data. Agreement was high when the SDB index was elevated and lower when the SDB index was in the mild-to-moderate range, regardless of the technique used to obtain it.

**Keywords:** Sleep disordered breathing, unattended limited monitoring, polysomnography, clinical history, agreement

**Citation:** Masdeu MJ; Ayappa I; Hwang D; Mooney AM; Rapoport DM. Impact of clinical assessment on use of data from unattended limited monitoring as opposed to full-in lab PSG in sleep disordered breathing. *J Clin Sleep Med* 2010;6(1):51-58.

The increasing use of unattended limited monitoring (ULM) in the evaluation of patients suspected of sleep disordered breathing (SDB) has prompted discussion of its impact on patient care. Published studies have shown generally good agreement when comparing SDB indices (AHI, RDI) obtained from several commercially available ULMs to those obtained from conventional full polysomnography (PSG).<sup>1-6</sup> However, it is assumed by most sleep clinicians that not only the value of the SDB index but the clinical algorithm plays a role in the diagnostic and therapeutic choices made by the clinician; furthermore, both the American Academy of Sleep Medicine (AASM) and Center for Medicare and Medicaid Services have issued documents that emphasize the need for incorporating comprehensive clinical assessment by a sleep specialist into diagnostic and therapeutic decisions.<sup>7</sup>

The purpose of the present study was to examine agreement among multiple sleep clinicians when presented with 2 versions of the usual clinical situation: clinical data plus the full attended PSG (including summary measures and the raw tracing), and identical clinical data plus only an unattended limited monitoring (including summary measures and the raw tracing) *in place of* the PSG.

## BRIEF SUMMARY

**Current Knowledge/Study Rationale:** Previous work has focused on the effect of using full nocturnal polysomnography versus portable limited monitoring on indices of sleep disordered breathing (SDB). This study examines the effect of using full nocturnal polysomnography vs. portable limited monitoring on diagnosis and treatment recommendations made by sleep clinicians when presented with clinical data along with the indices.

**Study Impact:** When presented with clinical data along with sleep data, sleep trained clinicians showed similar agreement on diagnoses and treatment recommendations whether the data was from full polysomnography or portable limited monitoring. Differences in diagnosis and treatment recommendations appeared to be driven by SDB severity; good agreement was found in patients with high SDB indices and poorer agreement in patients with low SDB indices, regardless of the method in which the index was obtained.

## METHODS

The data for the present study were obtained from subjects who participated in a previous validation study of an ULM device.<sup>6</sup> These patients were recruited from all those seen at the New York University (NYU) Sleep Disorders Center presenting with sleep complaints suggestive of SDB and willing

to undergo 2 nights of ULM followed by a full PSG (always in this order, as the original validation study had as one goal capturing the success rate of ULM use without prior sleep laboratory contact). Of note, only patients with a high likelihood of *obstructive* events were selected. Specifically, no patients with congestive heart failure or suspected central SDB were included because ULMs do not assess respiratory effort. In addition, to cover the full range of severity of SDB from normal to abnormal, we included volunteers recruited without regard to symptoms. Data collected on each subject included a full clinical history (Hx) focusing on chief complaint and symptoms, ULM during 2 nights at home with ARES Unicorder, and full in-laboratory attended PSG. History was collected at an initial intake, guided only by a loose outline, completed for each patient by a physician not part of the present study. A report of a full physical exam was also available on each patient. Complete data were obtained in 66 patients and 19 volunteers.

Excessive daytime somnolence (EDS) was defined as present in each subject if the Epworth Sleepiness Scale score was  $> 10$ ,<sup>8</sup> OR if the original interview records contained clear symptoms stated by the patient of fatigue, tiredness or sleepiness.

The ARES Unicorder is a self-applied limited monitoring device worn on the forehead. The signals recorded include oxygen saturation (SpO<sub>2</sub>) and pulse rate from reflectance forehead oximetry, airflow from a nasal cannula and pressure transducer, snoring levels via a calibrated acoustic microphone, head movement actigraphy and head position from accelerometers. The device also provides audible alerts during the study if poor quality airflow or SpO<sub>2</sub> is detected so that the subject can reposition the device. The AHI and RDI from this device have been compared to full PSG<sup>4,6</sup> and shown to have no systematic bias when done simultaneously.

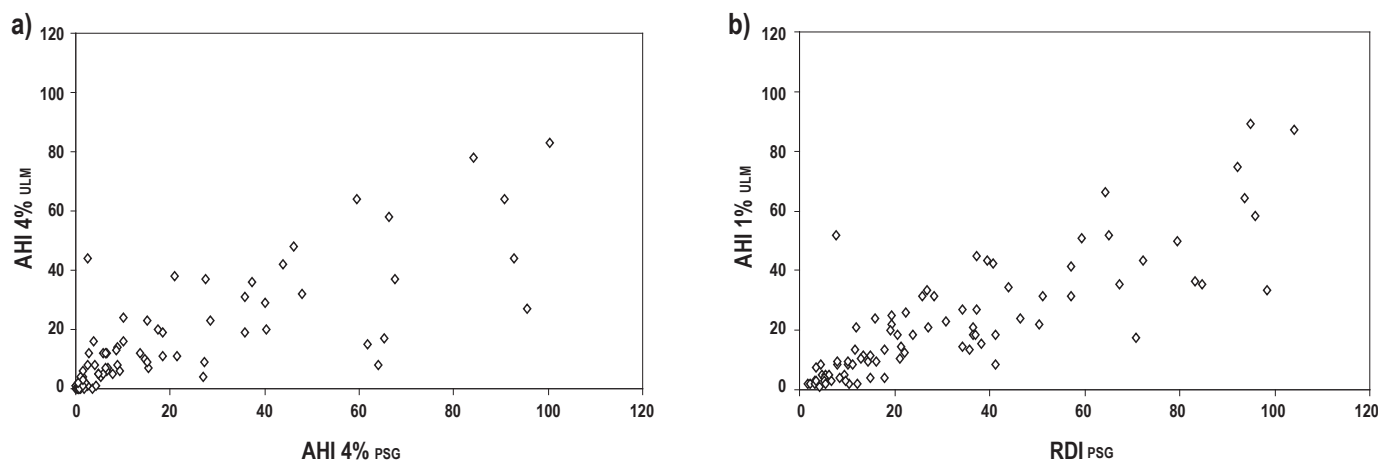
In-laboratory PSG was performed as per standard clinical guidelines<sup>9</sup> and included full electroencephalogram (EEG), electrooculogram, chin electromyogram, airflow recorded with a nasal cannula/pressure transducer and an oral thermistor, finger pulse oximetry, position monitor, anterior tibialis electromyogram and electrocardiogram.

For the PSG data, sleep was scored using Rechtschaffen and Kales criteria.<sup>10</sup> Arousals and periodic legs movements were scored by AASM standards.<sup>11</sup> Respiratory events were scored manually as follows: *Apneas* were identified when the airflow amplitude on the nasal cannula was  $< 10\%$  of baseline. *Hypopneas 4%* were identified when airflow amplitude was reduced by 30 % baseline and the event was followed by 4% oxygen desaturation. AHI4% was defined as the sum of apneas and hypopnea 4% divided by total sleep time. For calculation of the RDI, additional events were identified when airflow amplitude was  $< 50\%$  baseline or, alternatively, whenever a discernible change occurred in the airflow amplitude (generally amplitude was between 50% and 80% of the baseline) and the event was followed by 4% oxygen desaturation within 30 seconds, or an EEG arousal occurred within 5 seconds of its end. Thermistor and chest wall/abdominal movement detectors were only analyzed when the nasal cannula signal was uninterpretable, which occurred rarely in these studies ( $< 1\%$  of time), as technicians intervened immediately when the signal failed. All studies were scored by a single sleep technician.

As in previous work,<sup>6</sup> ARES Unicorder data was scored using an automated algorithm followed by technical review. Automated event analysis was performed using the ARES Insight software, with SDB events identified as follows: *Apnea* was identified when there was a cessation of airflow for 10 seconds. *Hypopnea 4%* included events identified as a hypopnea (airflow  $< 50\%$ ) with  $\geq 3.5\%$  desaturation followed by a 1% resaturation. In addition, *Hypopnea 1%* events were determined as those hypopneas with a minimum 1% desaturation and resaturation and at least one surrogate arousal indicator (head movement, change in snoring, or change in pulse rate). To calculate a valid recording time (denominator for SDB indices), the ARES automatically excluded (a) the first 15 min and last 3 min of the study, (b) periods in the record with poor or bad airflow or SpO<sub>2</sub> quality, (c) periods with excessive head movement, (d) any period when the patient was upright, and/or (e) the first 30 sec after each change of head position. After automated scoring, each ARES study was subjected to technical review by visual inspection of the signals, and SDB indices were calculated from the technician-edited data. AHI4%<sub>ULM</sub> was defined as apneas plus hypopneas 4% divided by valid time, and AHI1%<sub>ULM</sub> was apneas plus hypopneas 1% divided by valid time as this has been shown to correlate with the RDI.<sup>6</sup>

Two assessment packages were created for each subject. Each contained a copy of the identical full history (including all documents relating to the presence of EDS) and physical exam as described by a sleep clinician and the ARES symptom questionnaire filled out by the patient. Of note, the derived global assessment of sleepiness used to characterize patients in **Table 1b** was not specifically reported in this package. The ULM package included the report from the ARES analyses *and the electronic record* of the 2 nights of physiologic signals collected on the ARES Unicorder (Hx+ULM). The PSG package included the full PSG report *and the electronic record* of the signals collected on the PSG (Hx+PSG). The names of the subjects were not removed from the assessment packages given to the clinicians in the study (see discussion).

The clinicians were instructed to review the entire applicable electronic file of raw PSG or ULM data in each assessment. The assessment packages were presented separately to 4 sleep-trained clinicians working independently. All clinicians had experience and training in sleep medicine (2, 3, 6, 30 y). Three were board eligible at the time of the study, and the fourth clinician had been trained abroad but had been affiliated with our center for over 6 months. Each clinician reviewed each subject's 2 packages 2 weeks apart in random order. For each subject and each package the clinician was required to make a diagnosis and treatment recommendation. We limited the diagnoses to 5 categories: Normal (no sleep disorder); SDB; primary snoring; other sleep disorder (insomnia, insufficient sleep, restless legs syndrome, periodic leg movement disorder, medications, central nervous system hypersomnia); and inconclusive (despite review of sleep study and clinical data available). The treatment recommendations were limited to 4 categories: No Treatment; continuous positive airway pressure (CPAP) (or CPAP trial defined as therapeutic trial of CPAP to determine diagnosis based on response to therapy); Other treatment (upper airway surgery, dental appliance, positional therapy, sleep hygiene, drugs); and Repeat study (for reasons including insufficient time, insuffi-

**Figure 1**—Correlation between SDB indices obtained by the two recording techniques (PSG and ULM).a) Correlation between AH14%<sub>PSG</sub> vs. AH14%<sub>ULM</sub>. Each point represents one subject. b) Correlation between RDI<sub>PSG</sub> vs. AH14%<sub>ULM</sub>.

cient quality, insufficient time in supine position, need for EEG data if reviewing a ULM).

### Tabulation and Analysis

Correlations between SDB index (AH14% or RDI) obtained using ULM and PSG were examined using 2-way random effects intraclass correlation coefficients (ICC). Agreement, diagnostic sensitivity and specificity to SDB were calculated using a cut off of 15/h for RDI and 10/h for AH14%.

For agreement on diagnosis, comparisons were as follows: (i) between clinicians, all using Hx+PSG alone (ii) between clinicians, all using Hx+ULM (iii) within clinicians, using Hx+PSG vs. using Hx+ULM. For agreement on treatment recommendations, similar comparisons were tabulated. A summary of secondary diagnoses was also done when clinicians reviewed the clinical data.

All comparisons were performed separately on the following datasets and subsets:

- all 85 subjects (including the 66 patients and 19 non-patient volunteers)
- all 66 patients (excluding the 19 non-patient volunteers)
- all patients with RDI > 30/hr (n = 32)
- all patients with 10 < RDI < 30 (n = 25)
- all patients with RDI < 10/hr (n = 9)

All subjects signed a consent form approved by the institutional review board at the NYU School of Medicine.

## RESULTS

Data were analyzed on a total of 85 subjects (66 patients and 19 volunteers). There were 61 men and 24 women. **Table 1a** shows the demographic data of patients and volunteers. **Table 1b** shows the summary of symptoms for all subjects. The predominant complaint was snoring with/without EDS.

### Comparison of SDB Indices Using PSG and ULM

**Figure 1** shows the correlation between SDB indices obtained by the 2 recording techniques. **Figure 1a** shows the cor-

**Table 1a**—Demographic data

	Patients (n = 66, 51m /15f)			Volunteers (n = 19, 10m /9f)		
	Mean	Min	Max	Mean	Min	Max
Age, y	45	27	74	36	23	73
BMI, kg/m <sup>2</sup>	31	21	70	24	19	32
ESS	9	0	21	5	1	14
RDI/h on PSG	37	4	104	12	2	46
RDI/h on ULM	26	2	86	8	1	26

m, male; f, female; BMI, Body mass index; ESS, Epworth Sleepiness Scale; RDI/hr, Respiratory disturbance index per hour; PSG, Polysomnography; ULM, Unattended limited monitoring

**Table 1b**—Summary of symptoms

	Patients (n = 66)	Volunteers (n = 19)
Snoring alone	5	2
Snoring + definite* EDS	33	2
Snoring + questionable# EDS	27	1
EDS alone	1	2
None	0	12

EDS, Excessive Daytime Somnolence; \*Epworth Sleepiness Scale > 10 or clear symptoms of EDS including fatigue presented to clinician; #Epworth Sleepiness Scale < 10 and any symptoms of EDS including fatigue presented to clinician

relation between AH14%<sub>PSG</sub> vs. AH14%<sub>ULM</sub>. The ICC was 0.76. Using the physiologic data alone (not considering the clinician assessment), the sensitivity for diagnosis of SDB using a cut-off of AH14% > 10/h was 86% and specificity was 84%. Overall diagnostic agreement was 85%. **Figure 1b** shows the correlation

between RDI<sub>PSG</sub> vs. AHI1%<sub>ULM</sub>. The ICC was 0.8. The sensitivity for diagnosis of SDB using a RDI cutoff > 15/h was 81% and specificity was 94%. Overall diagnostic agreement (from numerical RDI alone) was 86%.

### Agreement between Clinicians on Diagnosis and Treatment Using Hx+PSG vs Hx+ULM

Figure 2 shows the comparison of diagnostic and treatment decisions between pairs of clinicians (6 combinations for the 4 clinicians). This was done independently for Hx+PSG and Hx+ULM. Data are presented separately for the entire group of subjects (n = 85) and for the restricted subgroup of patients only (n = 66).

For diagnosis, agreement between pairs of clinicians using Hx+PSG (panel a) ranged from 74% to 86% for the entire group and 74% to 89% for the subgroup of patients. Similarly, agreement between pairs of clinicians using Hx+ULM (panel b) ranged from 66% to 85% for the entire group and 65% to 85% for the subgroup of patients. The same pattern was evident in treatment decisions: agreement between pairs of clinicians using Hx+PSG ranged from 74% to 86% (panel c), and using Hx-ULM ranged from 58% to 77% (panel d).

### Agreement on Diagnosis and Treatment Using Hx+ULM vs Hx+PSG by Individuals

Comparison is shown in Figure 3a of the decisions on diagnosis made by individual clinicians when evaluating the package containing Hx+PSG to the package containing Hx+ULM. The overall agreement between decisions for each individual clinician was 77% (range 69% to 86%). Data were similar in the entire group of subjects and for the restricted subgroup of patients only.

Table 2 shows the pattern of agreement (diagonal) and disagreement (off-diagonal) between the diagnoses made by individual clinicians presented with Hx+PSG vs. Hx+ULM. As expected with the population presenting to our sleep disorders center, the majority of the patients were given a diagnosis of SDB or snoring. Only one subject was classified by only one clinician as “inconclusive” by both Hx+PSG and Hx+ULM. When using Hx+ULM, in 31 of 340 (9%) subject/clinician interactions the clinician gave the subject a diagnosis of “inconclusive”; of these, by Hx+PSG 18 were classified as SDB, 5 were “snoring,” and 6 were given other diagnoses, and 1 was classified as normal. Clinicians gave a diagnosis of “inconclusive” in 6 of 340 (2%) subject/clinician interactions when using Hx+PSG; by Hx+ULM, 4 of these were classified as SDB and 1 as “other diagnosis.” Thus the number of inconclusive diagnoses was higher (9% vs 2%) when the clinicians used the package Hx+ULM, especially in the group with RDI < 10/h.

Comparison is shown in Figure 3b of the decisions on treatment made by individual clinicians when evaluating packages containing either Hx+PSG or Hx+ULM. The overall agreement between decisions for each individual clinician was 75% (range 67% to 81%). Data were similar in the entire group of subjects and for the restricted subgroup of patients only.

### Agreement on Diagnosis and Treatment According to Severity of RDI

Figure 4 shows data for the patients alone, grouped by severity of RDI as measured by PSG. Diagnostic agreement between

**Table 2**—Cross-tabulation of agreement and disagreement on diagnoses by individual clinicians using Hx+PSG vs. Hx+ULM (n = 85 × 4)

Hx+ULM	Hx+PSG				
	Normal	SDB	Snoring	Inconclusive	Other
Normal	43	2	1	0	3
SDB	4	189	8	4	7
Snoring	1	8	15	0	1
Inconclusive	1	18	5	1	6
Other	0	4	3	1	15

SDB, Sleep disordered breathing; Hx+PSG, Clinical History plus polysomnography package; Hx+ULM, Clinical History plus unattended limited monitoring package

**Table 3**—Secondary diagnoses

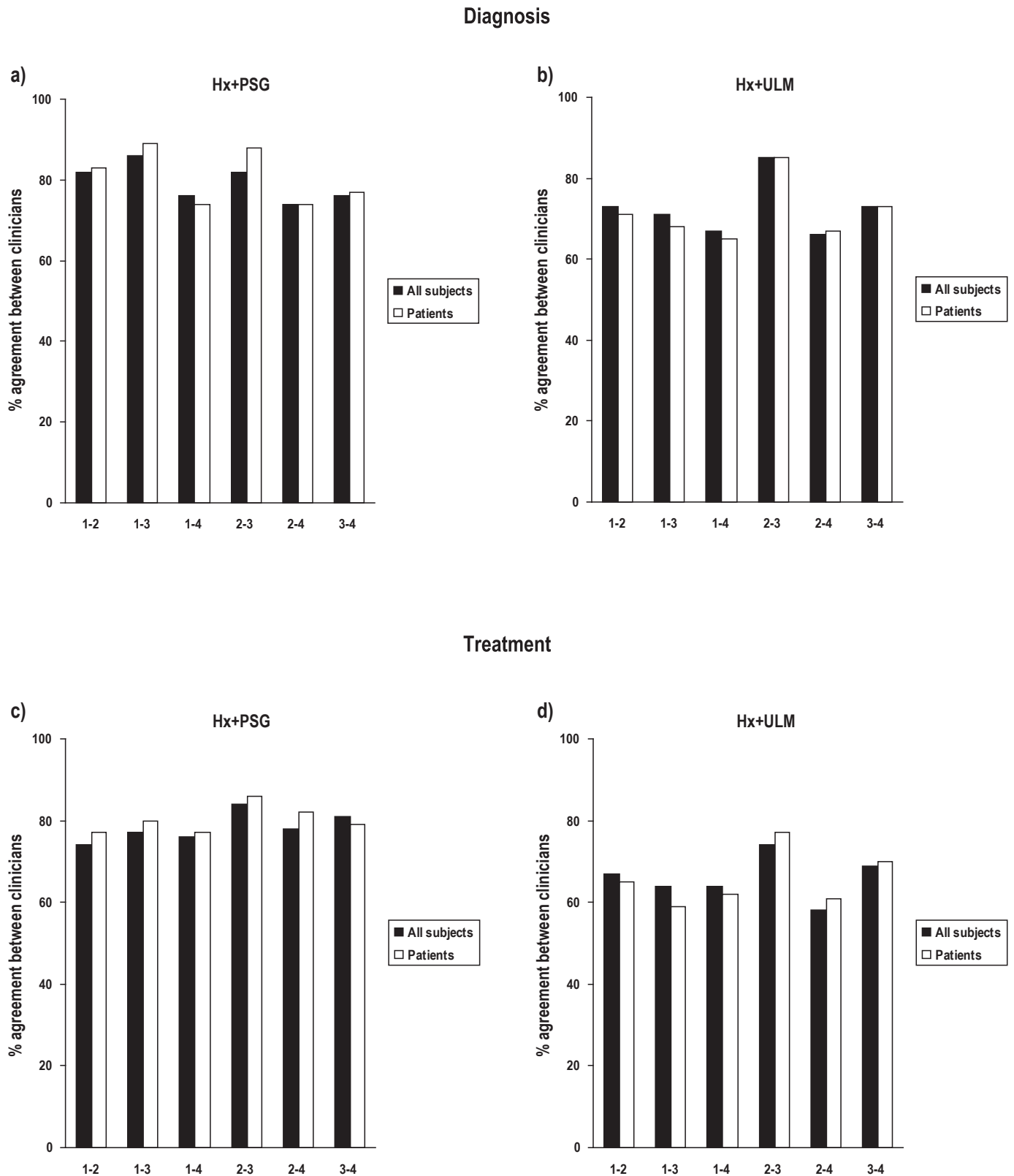
	Clinician 1		Clinician 2		Clinician 3		Clinician 4	
	PSG	ULM	PSG	ULM	PSG	ULM	PSG	ULM
Insomnia	1		2		3	3	5	3
Non respiratory hypersomnia	3	1			1	2	3	4
Insufficient sleep			1		3	2	1	2
PLMs			3		1		1	1
Other				2				

PSG, Polysomnography; ULM, Unattended limited monitoring; PLMs, Periodic leg movements

clinicians (Figure 4a) was highest in the subjects with the highest RDI (89% to 98%) and fell off markedly at the lowest RDI (39% to 52%), regardless of whether the clinicians used the package with the Hx+PSG or Hx+ULM (first 2 bars in each cluster). This pattern was also seen for the decisions made by an individual clinician using Hx+ULM vs. his/her decision made using Hx+PSG (third bar in each cluster).

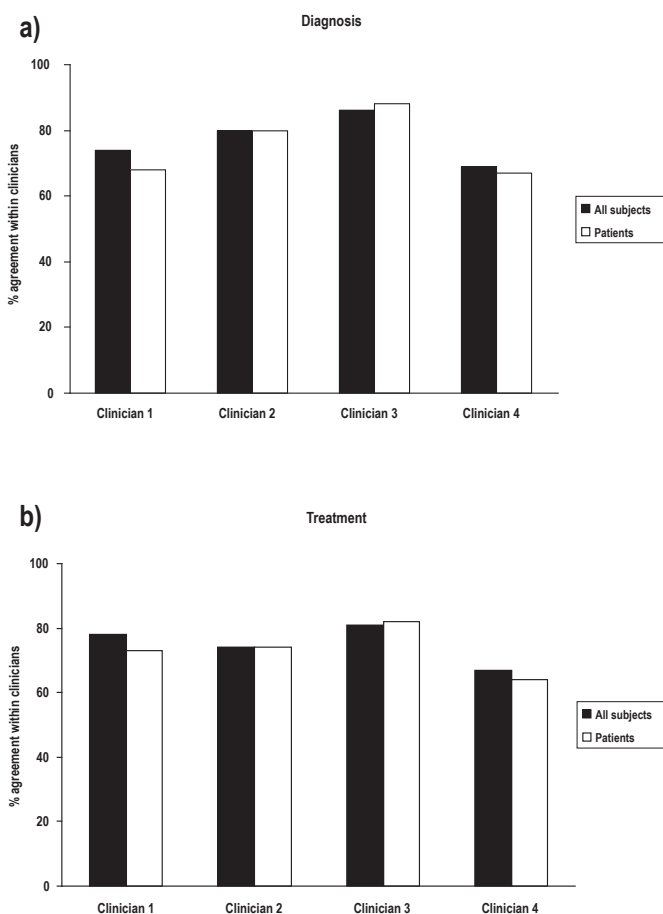
Agreement on treatment (Figure 4b) between clinicians was also high in the subjects with the highest RDI (84% to 97%) and lower in those with the lowest RDI (43% to 74%), irrespective of whether the clinicians used the package with the Hx+PSG or Hx+ULM (first 2 bars in each cluster). In patients with the lowest RDI (n = 9) the agreement between clinicians on treatment was substantially higher using Hx+PSG than Hx+ULM. The overall pattern of falling agreement with lower RDI was also seen when comparing decisions made by an individual clinician using Hx+ULM to decisions made using Hx+PSG (third bar in each cluster)

Table 3 shows a summary of secondary diagnoses using the 2 techniques. Similar secondary diagnoses were obtained when the information was primarily from the clinical information (e.g., insomnia, insufficient sleep). However, in 3 patients

**Figure 2**—Agreement between clinicians on diagnosis and treatment when using either Hx+PSG or Hx+ULM alone.

Black bar represents data for the entire group of subjects ( $n = 85$ ) and white bar represents data for the subgroup of patients ( $n = 66$ ). Comparisons are made between pairs of clinicians (6 combinations for the 4 clinicians). **a)** Agreement on diagnosis between pairs of clinicians using Hx+PSG. **b)** Agreement on diagnosis between pairs of clinicians using Hx+ULM. **c)** Agreement on treatment between pairs of clinicians using Hx+PSG. **d)** Agreement on treatment between pairs of clinicians using Hx+ULM.

**Figure 3**—Agreement on diagnosis and treatment when using Hx+ULM vs. Hx+PSG for each individual clinician.



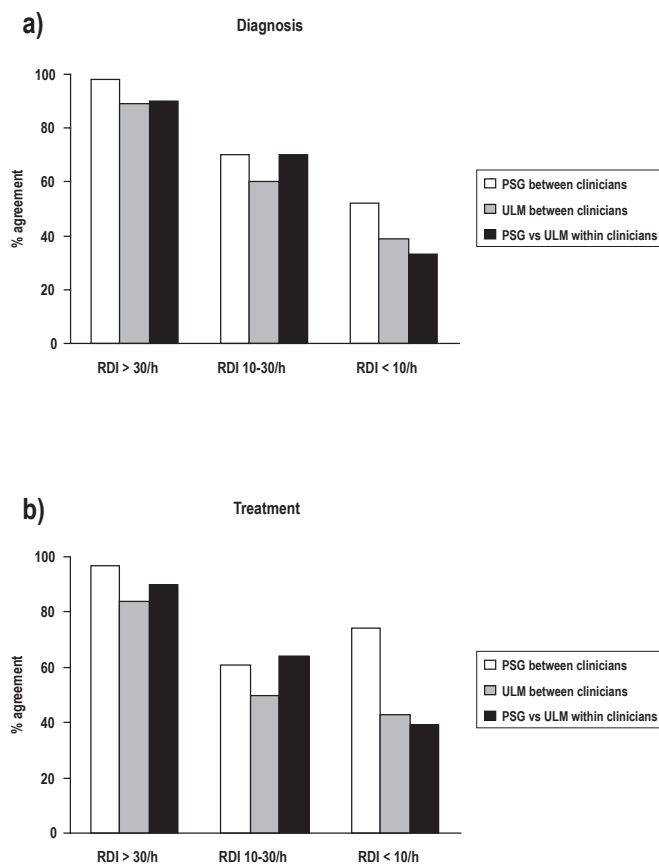
Black bar represents data for the entire group of subjects (n = 85) and white bar represents data for the subgroup of patients (n = 66). **a)** Comparison of the decisions on diagnosis made by individual clinicians when evaluating the package containing Hx+PSG to the package containing Hx+ULM. **b)** Comparison of the decisions on treatment made by individual clinicians when evaluating the package containing Hx+PSG to the package containing Hx+ULM.

a secondary diagnosis of periodic leg movement disorder was obtained using the PSG but was not available using the ULM.

## DISCUSSION

Our data show that sleep clinicians have significant disagreements for diagnosis and treatment recommendations even when presented with the “gold standard” of a PSG and clinical data. We are not aware of previous studies examining reliability of diagnoses in this manner when using the PSG. While agreement was high when the SDB index was elevated (irrespective of the technique used to obtain it), there was a much lower level of agreement (as low as 39%) when the SDB index was in the mild-to-moderate range. This was equally true of comparisons between clinicians using the same technique and between decisions of a single clinician using Hx plus the PSG vs. ULM.

**Figure 4**—Agreement on diagnosis and treatment for the subgroup of patients alone (n = 66), grouped by severity of RDI as measured on the PSG.



RDI > 30/h. (n = 32), RDI 10-30/h. (n = 25) and RDI < 10/h. (n = 9). First bar represents agreement between clinicians using Hx+PSG. Second bar represents agreement between clinicians using Hx+ULM. Third bar shows agreement for an individual clinician using Hx+PSG vs. Hx+ULM. **a)** Agreement on diagnosis. **b)** Agreement on treatment.

Analysis of our data allows some conclusions about the reasons for these results. When the disorder is severe (i.e., RDI > 30/h), clinicians do not differ significantly from each other when ULM or PSG are used to obtain the SDB index, perhaps because current accepted practice assumes that diagnosis and treatment for severe SDB is a straightforward clinical decision based on RDI. In contrast, when the SDB is in the mild-to-moderate range, and perhaps also when the clinical situation is more ambiguous (e.g., snoring without EDS, insomnia and insufficient sleep and a mid-range SDB index), clinicians often disagree with each other when using a single technique; they also make different decisions when presented with an SDB index obtained by PSG or ULM. There are several possible explanations for this situation:

1. Clinicians may disagree when presented with data from a single device due to giving different names to the same entity (e.g., *mild SDB* vs. *primary snoring*), or may have different thresholds for diagnosing individual syndromes (SDB index cutoff, symptoms to diagnose *insufficient sleep* and *insom-*

nia). This is supported by data from the comparison between clinicians using the PSG. **Figure 4** shows that clinicians may disagree less on treatment than they do on diagnosis in the group with the highest disagreement (RDI < 10/h) when using the PSG. However, this was not seen when clinicians used the ULM data. Disagreements in treatment recommendations in this subgroup (ULM) were predominantly caused by the requirement for a PSG/ multiple sleep latency tests by some clinicians but not by others, perhaps reflecting differences in the level of confidence in the data obtained from ULM.

- Individual clinicians may disagree with their own decisions when presented with Hx+ULM and Hx+PSG because there was actual night-to-night variation in the SDB index independent of the tool used to measure it.
- Individual clinicians presented with data from different devices may be using the additional information in one device over the other (e.g., the EEG data in full PSG), but this does not seem to occur in severe disease as much as in mild disease.
- Individual clinicians may be less willing to accept a low SDB index by ULM when this is discordant with a high clinical level of suspicion for SDB. Whether a PSG changes this is not directly addressed by our data, but our results are consistent with a higher level of “confidence” in the traditional metric, at least by some clinicians.

An assumption of the analysis used in the present study is that agreement among clinicians reflects the choice of the correct diagnosis or treatment. It can be argued that a more relevant metric for defining “correct diagnosis” might be the long-term clinical outcome experienced by the patient.

Further analysis of the studies labeled *inconclusive* by either ULM (31 of 340 instances) or PSG (6 of 340 instances) may be instructive. In 3 of 6 instances in which Hx+PSG resulted in an inconclusive decision, there was insufficient sleep time on the PSG. The remaining 3 subjects had RDI < 10/h on both studies, but the clinician apparently felt sufficient information was present to make a “diagnosis” from the analysis of the history and ULM package as evaluated at that time. Substantially more instances of inconclusive by Hx+ULM were present in our dataset. In 12 of 31 instances in which the diagnosis changed from inconclusive by ULM to SDB by PSG, this was associated with a low RDI on the ULM that increased on the PSG. This is consistent with the observation that RDI is often higher in the laboratory than in the home, seen in other studies,<sup>6,12</sup> as well as in the present data, mean  $\Delta(\text{RDI}_{\text{PSG}} - \text{RDI}_{\text{ULM}}) = 10 \pm 15.8$ . In 15 of 31 instances, the RDI on PSG and ULM were essentially identical and < 14/h. The diagnosis given with Hx+ULM was inconclusive, whereas the clinician chose a diagnosis after the same Hx plus the PSG. This represents a subset of cases where the PSG appears to confer greater information or greater security in the absence of disease to rule in or rule out diagnoses. The final 4 instances were cases in which the RDI was > 20/h on both PSG and ULM, but clinicians differed in how they related this to the same history.

A limitation of the present study lies in the patient population analyzed. Because we predominantly studied patients with a high risk of SDB, the agreement between a “negative ULM” (n = 41) and a “negative PSG” (n = 33), coupled to the clinical

history, has been only preliminarily examined and needs to be evaluated in a larger sample.

Another limitation of our study was that the analysis packages retained the names of the subjects and that memory of the prior reading could influence the reading of the second “package” for each patient. However, patients were not individually known to the clinicians by name, and presentation of the Hx+ULM vs Hx+PSG charts were separated by at least 2 weeks. While it would have been desirable to remove the names, we observed that the history itself provided more “identifying” details with potential for carryover than the name. We thus relied exclusively on randomizing and balancing the order in which the charts were presented to the 4 clinicians as well as separating the presentations in time. It should be noted that any memory of a patient from one reading to another would have improved agreement within clinicians, and thus our results may have understated disagreements.

Another limitation of this study is the method by which the clinicians were presented the data, specifically that they did not interact with the patients directly to obtain the detailed history and physical exam themselves. Unfortunately, it would have been impractical to have every patient seen by 4 clinicians (and the patients would have been biased by previous interactions had it been done in this way). Our choice of creating a paper chart was intended to expand on the previous comparisons of the PSG and ULM numerical data by performing at least a surrogate of a true patient interaction. However, it must be acknowledged that this may not be perfect.

Our demonstration that disagreements do occur among clinicians using the same data suggests additional data on the decision process may be needed. The focus need not remain exclusively on how one obtains the SDB index, but rather on which clinical situations require additional data (other than the metric of SDB).

There was a clear effect on the decisions made by clinicians of knowing which technique had been used to determine the SDB index. When the same decision was made by the clinician after (s)he was presented with the clinical information and either the RDI determined by the ULM or by the PSG, the RDI fell on the same side of the diagnostic cutoff value (i.e., both RDIs greater than or less than 15/h) in 89% of cases. This suggested a strong reliance on this absolute cutoff value in diagnosing SDB in this group of patients. However, even when a different decision (20% of cases) was made by the clinician using the 2 packets for a patient (i.e., clinical information and an RDI determined by either the ULM or the PSG), the RDI by the 2 techniques fell on the *same* side of the diagnostic cutoff value in 78% of these cases. Conversely, in the 12 cases in which the RDI using the 2 techniques fell on opposite sides of the diagnostic cutoff, the clinician nevertheless made the same diagnostic decision in 4 to 11 of 12 cases (mean 63%) when using all available data. In 5 of 12, the RDIs differed by < 10. This suggested that the clinicians may have used their opportunity to view the actual raw data and/or knowledge of the source of the RDI as part of their decision making algorithm.

In conclusion, our data show that when patients have severe SDB, analyzing data obtained by ULM or PSG produce similar high levels of agreement across clinicians and techniques. However, our data also show that when patients have mild SDB

and/or complex presentations (including insomnia, insufficient sleep, and other sleep disorders) agreement among clinicians is low for diagnosis no matter which index of SDB (ULM or PSG) is provided. This suggests that how the SDB index is obtained (ULM vs. PSG) is not the fundamental issue affecting their decision making. Future studies may need to continue to address the need for a full PSG, but our data suggest this need is dependent on the severity of the SDB, and data should be analyzed separately for different subgroups of patients.

## ABBREVIATIONS

ULM, Unattended limited monitoring  
PSG, Polysomnography  
SDB, Sleep disordered breathing  
Hx, Clinical history  
ARES Unicorder, Apnea Risk Evaluation System  
AHI, Apnea/hypopnea index  
RDI, Respiratory disturbance index  
NYU, New York University  
EDS, Excessive daytime somnolence  
SpO<sub>2</sub>, Oxygen saturation  
AASM, American Academy of Sleep Medicine  
EEG, Electroencephalogram  
CPAP, Continuous positive airway pressure  
ICC, Intraclass correlation coefficient

## REFERENCES

1. Ballester E, Solans M, Vila X, et al. Evaluation of a portable respiratory recording device for detecting apnoeas and hypopnoeas in subjects from a general population. *Eur Respir J* 2000;16:123-7.
2. Dingli K, Coleman EL, Vennelle M, et al. Evaluation of a portable device for diagnosing the sleep apnoea/hypopnoea syndrome. *Eur Respir J* 2003;21:253-9.
3. Ayappa I. Comparison of limited monitoring using a nasal-cannula flow signal to full polysomnography in sleep-disordered breathing. *Sleep* 2004;27:1171-9.
4. Westbrook PR. Description and validation of the apnea risk evaluation system: a novel method to diagnose sleep apnea-hypopnea in the home. *Chest* 2005;128:2166-75.
5. Yin M, Miyazaki S, Ishikawa K. Evaluation of type 3 portable monitoring in unattended home setting for suspected sleep apnea: factors that may affect its accuracy. *Otolaryngol Head Neck Surg* 2006;134:204-9.

6. Ayappa I, Norman RG, Seelall V, Rapoport DM. Validation of a self-applied unattended monitor for sleep disordered breathing. *J Clin Sleep Med* 2008;4:26-37.
7. Collop NA, Anderson WM, Boehlecke B, et al. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med* 2007;3:737-47.
8. Johns MW. Daytime sleepiness, snoring, and obstructive sleep apnea. The Epworth Sleepiness Scale. *Chest* 1993;103:30-6.
9. 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;22:667-89.
10. Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. University of California, Los Angeles. Brain Information Service; 1968.
11. Iber C. The AASM manual for the scoring of sleep and associated events: rules terminology and technical specifications. 1st ed. Westchester, IL: American Academy of Sleep Medicine; 2007.
12. Iber C, Redline S, Kaplan Gilpin AM, et al. Polysomnography performed in the unattended home versus the attended laboratory setting—Sleep Heart Health Study methodology. *Sleep* 2004;27:536-40.

## SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication January, 2009

Submitted in final revised form June, 2009

Accepted for publication June, 2009

Address correspondence to: David M. Rapoport, M.D., Sleep Disorders Center, 462 First Ave, NBV7N2, New York, NY 10016; Tel: (212) 263-6407; Fax: (212) 263-7445; E-mail: david.rapoport@nyu.med.edu

## DISCLOSURE STATEMENT

Research supported by grants from Advanced Brain Monitoring, NCRR M01RR00096 and Foundation for Research in Sleep Disorders. Dr. Rapoport has received research support from Fisher Paykel Healthcare, Medtronic, Ventus Medical, and Advanced Brain Monitoring and holds multiple US and foreign patents covering techniques and analysis algorithms for the diagnosis of OSAHS and techniques for administering CPAP. Several of these have been licensed to Biologics, Fisher Paykel Healthcare, Advanced Brain Monitoring, and Tyco. Dr Ayappa has received research support from Fisher Paykel Healthcare, Ventis Medical, and Advanced Brain Monitoring and holds multiple US and foreign patents covering techniques and analysis algorithms for the diagnosis of OSAHS and techniques for administering CPAP. Several of these have been licensed to Fisher Paykel Healthcare and Advanced Brain Monitoring. The other authors have indicated no financial conflicts of interest.