

# Polysomnography indexes are discordant with quality of life, symptoms, and reaction times in sleep apnea patients

EDWARD M. WEAVER, MD, MPH, B. TUCKER WOODSON, MD, and DAVID L. STEWARD, MD, Seattle, Washington, Milwaukee, Wisconsin, and Cincinnati, Ohio

**OBJECTIVE:** We tested whether polysomnography (PSG) indexes were associated with sleepiness, quality of life, or reaction times at baseline and as outcome measures following surgical or sham treatment for patients with obstructive sleep apnea syndrome (OSAS).

**STUDY DESIGN AND METHODS:** Mild-moderate OSAS subjects were measured before and 8 weeks after surgical or sham treatment in this prospective longitudinal study. Measures included standard PSG indexes, sleepiness, quality of life, and reaction times. Associations were examined with Spearman correlations and multivariate linear regression.

**RESULTS:** Correlations between baseline PSG and non-PSG measures ranged from  $-0.22$  to  $0.25$  ( $n$ , 87 subjects; mean correlation,  $0.00 \pm 0.11$ ), with one positive association significant of 56 tested (arousal index and SF36 Mental Component Summary,  $r$ ,  $0.25$ ;  $P = 0.03$ ). Correlations between change in PSG and non-PSG measures ranged from  $-0.37$  to  $0.35$  ( $n$ , 54 subjects; mean correlation,  $-0.05 \pm 0.19$ ), with no significant positive association of 56 tested. Regression analyses confirmed these results.

**CONCLUSIONS:** PSG indexes are not consistently associated with sleepiness, quality of life, or reaction time, both at baseline and as outcome measures in patients with mild-moderate OSAS. PSG in-

dexes may not quantify some important aspects of OSAS disease burden or treatment outcome. Clinically important outcomes should be measured directly. **EBM rating: A.** (Otolaryngol Head Neck Surg 2005;132:255-62.)

**P**olysomnography (PSG) is considered the gold standard for the diagnosis of obstructive sleep apnea syndrome (OSAS), the estimation of its severity, and the measurement of treatment response.<sup>1</sup> Although OSAS includes both a PSG abnormality and symptoms,<sup>2</sup> its severity is often defined by the apnea-hypopnea index (AHI) alone.<sup>2</sup> Improvement of symptoms, quality of life, and function are important outcomes, especially for patients with milder OSAS where serious medical complications are less likely to occur.<sup>3-6</sup>

Surgical treatments for OSAS have typically been judged on PSG outcomes. Anecdotally, however, patients' reports of improvement often are discordant with PSG outcomes. If true, then clinically important outcomes like symptoms, quality of life, and reaction times should be measured directly to evaluate surgical (and nonsurgical) treatment response.

Previous reports indicate a poor association at baseline between standard PSG indexes and sleepiness, health status, and quality of life in different populations of OSAS patients.<sup>7,8</sup> We sought to determine whether PSG indexes are associated with sleepiness, quality of life, or reaction times at baseline and as outcomes after active or sham treatment, in a new sample of patients with mild to moderate OSAS.

## METHODS

### Study Design

A subgroup analysis of data from a 2-institution randomized controlled trial<sup>9</sup> was performed to test the hypotheses that PSG indexes are not associated with sleepiness, quality of life, and reaction times before treatment or as outcomes 8 weeks after surgical or sham treatment for mild to moderate OSAS.

### Participants

Eligible subjects were adults with (1) excessive daytime sleepiness, (2) mild or moderate OSAS on screening sleep study, (3) no prior treatment for OSAS, and (4) no morbid obesity. Detailed inclusion and exclusion

---

From the Department of Otolaryngology-Head and Neck Surgery (Dr Weaver), Sleep Disorders Center (Dr Weaver), and Center for Cost and Outcomes Research (Dr Weaver), University of Washington, Seattle, WA; the Department of Otolaryngology and Communication Sciences (Dr Woodson) and Sleep Disorders Program (Dr Woodson), Medical College of Wisconsin, Milwaukee, WI; and the Department of Otolaryngology-Head and Neck Surgery (Dr Steward), University of Cincinnati, Cincinnati, OH.

Data collection was supported in part by a research grant from Gyrus-ENT. Dr. Weaver was supported by the Robert Wood Johnson Clinical Scholars Program during project development.

Dr. Weaver is currently supported by a career development award (HL068849) from the National Heart, Lung, and Blood Institute, and by a career development scholars award from the American Geriatrics Society.

Presented at the Annual Meeting of the American Academy of Otolaryngology-Head and Neck Surgery, Orlando, FL, Sept. 21-24, 2003.

Reprint requests: Edward M. Weaver, 1660 S. Columbian Way (112-OTO), Seattle, WA 98108; e-mail, eweaver@u.washington.edu.

0194-5998/\$30.00

Copyright © 2005 by the American Academy of Otolaryngology-Head and Neck Surgery Foundation, Inc.

doi:10.1016/j.otohns.2004.11.001

criteria were described previously.<sup>9</sup> Age, sex, and body mass index were recorded for all subjects at baseline. All available baseline and outcomes data were analyzed from the parent study.

Subjects were recruited directly from 2 academic otolaryngology practices and from poster and newspaper advertisements. This study was designed and performed with approval from local institutional review boards. All patients gave informed consent.

### Polysomnography

All subjects underwent an overnight, monitored, in-laboratory diagnostic PSG at baseline (separate from the screening sleep study) and 8 weeks after completion of surgical or sham treatment. PSG measurements included an electroencephalogram ( $\geq 2$  channels), electrooculogram, chin and leg muscle electromyograms, electrocardiogram, measures of oronasal airflow, thoracic and abdominal efforts, body position, and pulse oximetry.

Sleep stages<sup>10</sup> and arousals<sup>11</sup> were scored in standard fashion. Apnea was defined as cessation of inspiratory airflow  $\geq 10$  seconds. Hypopnea was defined as a reduction of inspiratory airflow  $\geq 10$  seconds with an associated 4% decrease in oxyhemoglobin saturation or an electroencephalographic arousal. Respiratory arousals were quantified on PSGs at one study site and were defined as electroencephalographic arousals associated with an apnea or hypopnea. All PSGs were manually scored by trained technicians and confirmed by a blinded, board-certified sleep physician.

### Intervention

Of 90 subjects enrolled in the parent trial, 29 received sham radiofrequency tongue reduction, 28 received radiofrequency tongue and palate reduction, and 27 received continuous positive airway pressure therapy.<sup>9</sup> For this study, we analyzed baseline data for all subjects and outcomes data for the subjects who received sham radiofrequency tongue reduction or active radiofrequency tongue and palate reduction. Continuous positive airway pressure subjects did not undergo post-treatment PSG and thus data were not available to analyze the association between PSG outcomes and other outcomes in these subjects. The sham radiofrequency tongue reduction and the active radiofrequency tongue and palate reduction have been described in detail.<sup>9</sup> Active treatment included a mean total  $9700 \pm 2000$  J of energy to the tongue over 5 treatment sessions and  $1785 \pm 904$  J of energy to the palate over 2 treatment sessions.<sup>9</sup>

### Polysomnography Indexes and Non-PSG Measures

AHI, apnea index, total arousal index, and respiratory arousal index were calculated as the number of events, respectively, per hour of sleep. Lowest saturation represented the nadir oxyhemoglobin saturation. The percent sleep time with a saturation below 90% and percent sleep time in rapid eye movement stage of sleep (%REM) were calculated.

Non-PSG measures were chosen to represent meaningful measurements of quality of life, sleepiness, and reaction times. OSAS-specific quality of life was measured with 2 validated questionnaires: (1) Functional Outcomes of Sleep Questionnaire,<sup>12</sup> and (2) Symptoms of Nocturnal Obstruction and Related Events, formerly the OSA Patient Oriented Severity Index.<sup>13</sup> General health status was measured with the SF-36 (version 1) Mental Component Summary and Physical Component Summary Scales.<sup>14</sup> Daytime sleepiness was measured using the Epworth Sleepiness Scale.<sup>15</sup> Reaction times were measured using the Psychomotor Vigilance Task (PVT-192, Ambulatory Monitoring Inc, Ardsley, NY) with a total test time of 10 minutes and stimulus interval between 2 and 10 seconds.<sup>16</sup> Slowest reaction time was measured as the mean of the slowest 10% of reaction times. Slowest reaction time was analyzed as the reciprocal ( $1/\text{slowest reaction time}$ ) to minimize the contribution of very long lapses.<sup>16</sup> Median reaction time and fastest reaction time (mean of fastest 10% reaction times) were also analyzed.

PSG and non-PSG measures were compared at baseline. Change in PSG and non-PSG measures were compared as outcomes. Change was calculated as the difference between the baseline and the 8-week posttreatment measurement, with a positive value denoting improvement and a negative value denoting worsening.

Outcomes were measured at 8 weeks to allow time for complete recovery from the intervention (estimated 3 weeks)<sup>9</sup> and time to acclimate to improved sleep. Although longer-term outcomes would also be desirable, ethical considerations related to the delayed treatment in sham patients limited the parent study to 8-week outcomes.<sup>9</sup>

### Statistical Methods

All variables were analyzed as continuous variables. Data are presented as the mean  $\pm$  standard deviation. Nonnormal variables must be analyzed with nonparametric statistical tests, so normality was tested on all outcome variables. Normality was tested with the Shapiro-Wilk W test, Shapiro-Francia W' test, and com-

bined skewness and kurtosis tests. A variable was considered non-normal if it failed any one of these tests.

Spearman correlations were calculated between each PSG and non-PSG measure at baseline, and between change in each PSG and non-PSG measure as outcomes. Spearman correlations are presented because the assumptions of normal distributions for Pearson correlations were not met in many cases. A positive correlation denotes an expected correlation (eg, improvement on both variables or worsening on both variables). A negative correlation denotes an unexpected correlation (eg, improvement on one variable and worsening on the other variable). We sought to determine the positive correlation between PSG indexes and non-PSG measures. With  $N = 87$  baseline subjects, this study has  $>80\%$  power to detect an important baseline correlation coefficient ( $\geq 0.3$ )<sup>17</sup> at the 2-tailed significance level of 0.05.<sup>18</sup> With  $N = 54$  outcomes subjects, this study has 60% power to detect an important outcome correlation coefficient.<sup>18</sup>

Correlation analyses do not adjust for potential confounding variables. Therefore, to further assess for significant associations that may not be detected by simple correlations, we used multivariate linear regression. Associations were assessed between each non-PSG measure as the dependent (continuous) variable and certain PSG indexes as the independent (continuous) variable, adjusting for age, sex, and body mass index. AHI was analyzed in each case as was each PSG index that had a positive correlation  $\geq 0.20$ , even if not significant. Multiple linear regression is valid if the associations are linear and the variances are homogeneous (homoscedasticity). We tested these assumptions by visually analyzing residual diagnostics. When these assumptions were not met, we used normal transformations of the variables to meet these assumptions. Each adjustment variable is a potential confounder because each may be associated with PSG and non-PSG measures. 95% confidence intervals and  $P$  values (based on the coefficient  $t$  test) were calculated for the coefficient of each PSG measure tested.

All data were analyzed with Stata/SE 8.0 software (Stata Corp, College Station, TX). A  $P$  value  $< 0.05$  was considered statistically significant.

## RESULTS

The sample population was middle-aged, predominantly male, and overweight (Table 1). At baseline, the study sample had mild to moderate OSAS based on AHI and mild abnormality based on other PSG parameters (Table 2). The subjects had mild abnormalities in quality of life, sleepiness, and slowest reaction time (Table 3) at baseline. The sample of treated and sham subjects together showed a minimal mean change in

**Table 1.** Sample characteristics

Variable	Mean $\pm$ SD	Range
Baseline Sample (N = 87)		
Age (y)	49 $\pm$ 9	27-64
Sex (% male)	78	
Body Mass Index (kg/m <sup>2</sup> )	28 $\pm$ 4	19-37
Outcome Sample* (N = 54)		
Age (y)	48 $\pm$ 9	27-63
Sex (% male)	78	
Body Mass Index (kg/m <sup>2</sup> )	28 $\pm$ 4	22-37

SD, Standard deviation.

\*Outcomes sample includes subjects who received active treatment or sham treatment and underwent post-treatment polysomnography.

**Table 2.** Polysomnography summary

Variable	Mean $\pm$ SD	Range
Baseline (N = 87)		
Apnea-Hypopnea Index (events/h)	19 $\pm$ 10	2-44
Apnea Index (events/h)	6 $\pm$ 8	0-37
Arousal Index (events/h)	23 $\pm$ 10	0-51
Respiratory Arousal Index (events/h)	14 $\pm$ 9	0-32
Lowest Saturation (%)	87 $\pm$ 6	58-97
Saturation $<$ 90% (% of sleep time)	1.3 $\pm$ 2.1	0-8.5
REM Sleep (% of sleep time)	18 $\pm$ 6	4-32
Outcome* (N = 54)		
$\Delta$ Apnea-Hypopnea Index (events/h)	3 $\pm$ 12	-25-36
$\Delta$ Apnea Index (events/h)	0.2 $\pm$ 7.7	-17-31
$\Delta$ Arousal Index (events/h)	0.5 $\pm$ 11.6	-25-31
$\Delta$ Respiratory Arousal Index (events/h)	4 $\pm$ 11	-15-30
$\Delta$ Lowest Saturation (%)	-0.1 $\pm$ 4.6	-12-13
$\Delta$ Saturation $<$ 90% (% of sleep time)	-0.1 $\pm$ 2.2	-9-7
$\Delta$ REM Sleep (% of sleep time)	-2 $\pm$ 6	-18-12

SD, Standard deviation;  $\Delta$ , change from baseline (positive value denotes improvement, negative value denotes worsening).

\*Outcomes sample includes subjects who received active treatment or sham treatment and underwent post-treatment polysomnography.

PSG measures (Table 2) and a consistent mild mean improvement in all non-PSG measures (Table 3).

At baseline, of the 56 correlations examined between PSG and non-PSG measures, there were 25 (45%) positive correlations with just 1 (2%) statistically significant. There was 1 (2%) correlation  $\geq 0.20$  (Table 4). The mean correlation between PSG and non-PSG measures was  $0.00 \pm 0.11$ . In contrast, of 21 correlations between PSG measures, there were 19 (90%) positive correlations with 11 (52%) statistically significant. There were 12 (57%) correlations  $\geq 0.20$  (data not shown). The mean correlation between PSG measures was  $0.30 \pm 0.27$ . Of 28 correlations between non-PSG measures, there were 25 (89%) positive correlations with 11 (39%) statistically significant. There were 12 (43%) correlations  $\geq 0.20$  (data not shown). The mean correlation between non-PSG measures was  $0.26 \pm 0.27$ .

**Table 3.** Non-polysomnography variable summary

Variable	Mean $\pm$ SD Range	
Baseline (N = 85)		
Functional Outcomes of Sleep		
Questionnaire	16 $\pm$ 2	8-20
SNORE25	1.6 $\pm$ 0.7	0.4-3.4
SF-36 Mental Health Component Summary	47 $\pm$ 10	19-62
SF-36 Physical Health Component		
Summary	50 $\pm$ 8	26-63
Epworth Sleepiness Scale	12 $\pm$ 4	1-22
1/Slowest Reaction Time (1/msec)	2.8 $\pm$ 0.6	1.0-4.3
Median Reaction Time (msec)	230 $\pm$ 37	171-352
Fastest Reaction Time (msec)	186 $\pm$ 24	148-283
Outcome* (N = 54)		
$\Delta$ Functional Outcomes of Sleep		
Questionnaire	0.8 $\pm$ 2.1	-4.9-6.4
$\Delta$ SNORE25	0.3 $\pm$ 0.6	-0.8-1.5
$\Delta$ SF-36 Mental Health Component		
Summary	1.5 $\pm$ 6.9	-11-22
$\Delta$ SF-36 Physical Health Component		
Summary	1.0 $\pm$ 7.3	-11-29
$\Delta$ Epworth Sleepiness Scale	1.5 $\pm$ 3.5	-5-12
$\Delta$ 1/Slowest Reaction Time (1/msec)	0.2 $\pm$ 0.6	-1.2-1.9
$\Delta$ Median Reaction Time (msec)	7 $\pm$ 21	-47-52
$\Delta$ Fastest Reaction Time (msec)	6 $\pm$ 19	-46-101

SD, Standard deviation; SNORE25, Symptoms of Nocturnal Obstruction and Related Events questionnaire; SF-36, Medical Outcomes Trust Short Form 36 questionnaire;  $\Delta$ , change from baseline (positive value denotes improvement, negative value denotes worsening).

\*Outcomes sample includes subjects who received active treatment or sham treatment and underwent post-treatment polysomnography.

The only baseline correlation  $\geq 0.20$  between PSG and non-PSG measures occurred between baseline arousal index and SF36 Mental Component Summary scale, with a correlation 0.25 ( $P = 0.03$ ). However, on multivariate linear regression, this association was not significant ( $P = 0.57$ ) after adjusting for age, sex, and body mass index.

At outcome, of the 56 correlations examined between change in PSG and non-PSG measures, there were 21 (38%) positive correlations with none (0%) statistically significant. There were 8 (14%) correlations  $\geq 0.20$  (Table 5). The mean correlation was  $-0.05 \pm 0.19$ . In contrast, of 21 correlations between change in PSG measures, there were 15 (71%) positive correlations with 9 (43%) statistically significant. There were 14 (67%) correlations  $\geq 0.20$  (data not shown). The mean correlation between change in PSG measures was  $0.28 \pm 0.32$ . Of 28 correlations between change in non-PSG measures, there were 25 (89%) positive correlations with 8 (29%) statistically significant. There were 13 (46%) correlations  $\geq 0.20$  (data not shown). The mean correlation between change in non-PSG measures was  $0.24 \pm 0.23$ .

Multivariate linear regression analysis was used to examine the relationship between each of the 8 corre-

lations  $\geq 0.20$  between the change in PSG and non-PSG measures, adjusting for age, sex, and body mass index. With multivariate linear regression, the only significant association was between change in apnea index and change in SF36 Mental Component Summary ( $\beta = -0.36$ ; 95% confidence interval  $[-0.59, -0.12]$ ;  $P = 0.003$ ). This regression indicates that for every increase in apnea index by 10 events per hour, there is a decrease in the SF36 Mental Component Summary by 3.6, when age, sex, and body mass index are held constant. Two outliers drove this association. These outliers had large reduction in apnea index with large increases in hypopnea index (apneas converted to hypopneas). With removal of these outliers, the association is not significant ( $\beta = -0.20$ ; 95% confidence interval  $[-0.61, 0.21]$ ;  $P = 0.33$ ); however, we believe these outliers represent real results and should be left in the analysis.

The association between AHI and each of the non-PSG measures was analyzed with multivariate linear regression at baseline and at outcome, adjusting for age, sex, and body mass index. There was no significant association between baseline AHI and any baseline non-PSG measure nor between change in AHI and change in any non-PSG measure (Table 6). As examples, Figs 1 and 2 depict the poor association between Functional Outcomes of Sleep Questionnaire and AHI both at baseline and at outcome.

## DISCUSSION

Standard baseline PSG indexes correlate poorly with other measures of OSAS burden. These results are consistent with prior studies showing poor pre-treatment correlation between PSG and other measures in distinct clinical samples,<sup>7,8</sup> and this study extends these results to include objective measures of function (reaction times). Moreover, the changes in PSG indices correlate poorly with changes in other measures important to OSAS, including subjective sleepiness, OSAS-related quality of life, general health status, and function (reaction times). These data do suggest, however, that there is slightly greater association between some PSG and non-PSG outcomes than between PSG and non-PSG baseline measures. AHI, the primary PSG measure of OSAS severity and outcome, was not associated with any of the non-PSG measures at baseline or at outcome, even on adjusted analyses. These results suggest that the effects of OSAS and its treatment on clinically important outcomes should be measured directly and should not be inferred from surrogate PSG measures.

This type of discrepancy between physiologic and other measures of disease severity and outcome is not unique to OSAS. Investigators have raised similar is-

**Table 4.** Correlation\* table for baseline polysomnography and non-polysomnography measures

		PSG measures						
		AHI	AI	Arl	RArl	LSAT	%T < 90%	REM%
Non-PSG measures	FOSQ	0.07	0.01	0.11	-0.22	-0.04	0.01	0.02
	SNORE25	-0.06	-0.12	0.16	0.02	-0.15	-0.05	-0.04
	MCS	-0.01	-0.06	0.25†	-0.03	-0.13	0.00	0.15
	PCS	-0.11	-0.05	0.01	-0.20	0.12	0.09	-0.05
	ESS	-0.01	-0.01	0.12	0.06	-0.01	0.08	-0.09
	1/SRT	0.12	0.00	0.16	0.16	-0.21	-0.20	-0.07
	RT	0.10	-0.01	-0.03	0.14	-0.15	-0.15	0.03
FRT	0.19	0.03	0.00	0.18	-0.05	-0.08	0.05	

PSG, Polysomnography; AHI, Apnea Hypopnea Index; AI, Apnea Index; ArI, Arousal Index; RArl, Respiratory Arousal Index; LSAT, lowest oxyhemoglobin saturation; %T < 90, percent of sleep time with an oxyhemoglobin saturation of <90%; REM%, percent of sleep time in rapid eye movement sleep; FOSQ, Functional Outcomes of Sleep Questionnaire; SNORE25, Symptoms of Nocturnal Obstruction and Related Events questionnaire; MCS, SF-36 Mental Component Summary scale; PCS, SF-36 Physical Component Summary scale; ESS, Epworth Sleepiness Scale; 1/SRT, reciprocal of slowest reaction time; RT, median reaction time; FRT, fastest reaction time.

\*Positive correlation denotes expected correlation (e.g., both measures abnormal); negative correlation denotes unexpected correlation (e.g., one measure normal and the other abnormal).

†P < 0.05.

**Table 5.** Correlation\* table for change in polysomnography and non-polysomnography measures

		Δ PSG measures						
		AHI	AI	Arl	RArl	LSAT	%T < 90%	REM%
Δ Non-PSG measures	FOSQ	-0.09	-0.29†	-0.03	-0.24	-0.29†	-0.20	-0.21
	SNORE25	-0.04	-0.27	-0.13	0.16	-0.37†	-0.19	-0.11
	MCS	-0.16	0.24	-0.19	0.14	0.15	0.23	0.03
	PCS	0.05	-0.26	0.07	0.21	-0.31†	-0.22	-0.18
	ESS	0.03	-0.20	0.01	0.25	-0.12	-0.11	-0.16
	1/SRT	0.18	0.07	-0.04	0.10	-0.26	-0.13	-0.23
	RT	0.21	-0.04	-0.01	0.35	-0.29†	-0.08	-0.04
FRT	0.20	0.00	0.08	0.32	-0.37†	-0.22	0.03	

Δ, Change from baseline; PSG, polysomnography; AHI, Apnea Hypopnea Index; AI, Apnea Index; ArI, Arousal Index; RArl, Respiratory Arousal Index; LSAT, lowest oxyhemoglobin saturation; %T < 90, percent of sleep time with an oxyhemoglobin saturation of < 90%; REM%, percent of sleep time in rapid eye movement sleep; FOSQ, Functional Outcomes of Sleep Questionnaire; SNORE25, Symptoms of Nocturnal Obstruction and Related Events questionnaire; MCS, SF-36 Mental Component Summary scale; PCS, SF-36 Physical Component Summary scale; ESS, Epworth Sleepiness Scale; 1/SRT, reciprocal of slowest reaction time; RT, median reaction time; FRT, fastest reaction time.

\*Positive correlation denotes expected association (i.e., both measures improved or worsened); negative correlation denotes unexpected association (i.e., one measure improved and the other worsened, or vice versa).

†P < 0.05. Note: all significant correlations are negative.

sues with sinusitis, where there is no consensus on how to define severity of disease and outcomes of treatment, because subjective symptoms do not match objective findings on CT scan.<sup>19</sup> Similarly, objective pulmonary function tests do not correlate with symptoms in asthma patients,<sup>20</sup> and objective cardiac function tests do not correlate with quality of life in angina patients.<sup>21</sup>

These results do not mean that sleepiness, OSAS-related quality of life, and reaction times are irrelevant to OSAS. To the contrary, these measures help define OSAS.<sup>2</sup> Furthermore, the OSAS-specific quality of life measures (Functional Outcomes of Sleep Questionnaire; Symptoms of Nocturnal Obstruction and Related Events questionnaire) were derived from complaints by

OSAS patients.<sup>12,13</sup> The Psychomotor Vigilance Task used to measure reaction times is a validated objective measure of disorders of excessive sleepiness, including OSAS. Thus, this study suggests that PSG measures do not capture all elements of OSAS and thus should not be used exclusively to evaluate treatment response. It should be noted that patients rarely present for care in order to improve their AHI or other PSG parameter per se, but rather they present to improve symptoms, quality of life, and performance.

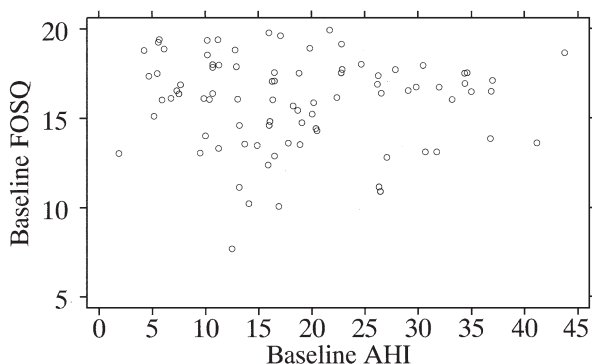
These comparisons of PSG and non-PSG outcomes have rarely been explicitly tested previously in an OSAS population. Walker-Engstrom et al<sup>22</sup> tested baseline and change correlations between PSG vari-

**Table 6.** Summary of multivariate linear regression models with apnea-hypopnea index as the independent variable

Dependent variable	Adjusted coefficient*	95% CI	P value
Baseline			
Functional Outcomes of Sleep Questionnaire	0.03	(-0.03, 0.08)	0.35
SNORE25	0.00	(-0.02, 0.01)	0.83
SF36 Mental Component Summary	0.06	(-0.16, 0.29)	0.58
SF36 Physical Component Summary	-0.05	(-0.22, 0.11)	0.53
Epworth Sleepiness Scale	0.02	(-0.09, 0.12)	0.74
1/Slowest reaction time	0.01	(-0.01, 0.02)	0.47
Median reaction time	0.37	(-0.47, 1.21)	0.38
Fastest reaction time	0.40	(-0.15, 0.96)	0.15
Outcome			
Δ Functional Outcomes of Sleep Questionnaire	-0.01	(-0.07, 0.05)	0.74
Δ SNORE25	0.00	(-0.01, 0.02)	0.82
Δ SF36 Mental Component Summary	-0.13	(-0.30, 0.05)	0.15
Δ SF36 Physical Component Summary	0.15	(-0.18, 0.21)	0.88
Δ Epworth Sleepiness Scale	0.02	(-0.06, 0.11)	0.54
Δ 1/Slowest reaction time	0.01	(-0.01, 0.02)	0.37
Δ Median reaction time	0.32	(-0.27, 0.91)	0.28
Δ Fastest reaction time	0.12	(-0.41, 0.64)	0.65

95% CI, 95% confidence interval; SNORE25, Symptoms of Nocturnal Obstruction and Related Events questionnaire.

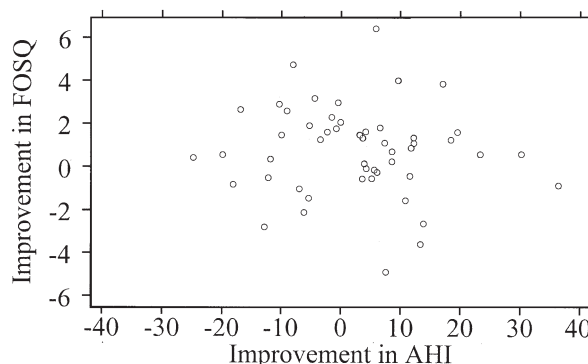
\*Coefficient describes the worsening in each dependent variable per unit increase (worsening) in apnea-hypopnea index, adjusting for age, sex, and body mass index.



**Fig 1.** Scatterplot of Functional Outcome of Sleep Questionnaire (FOSQ) and apnea-hypopnea index (AHI) at baseline ( $n = 86$ ). Note the apparent lack of association between these variables.

ables (AHI, apnea index, and oxygen desaturation index) and quality of life in their trial of dental appliance and uvulopalatopharyngoplasty patients. They found no significant correlations. Our study expands on these data by using OSAS-specific quality of life instruments and several other non-PSG measures as well as several more PSG indexes.

Furthermore, we examined potential associations further with multivariate regression to account for potential confounding variables. In addition, Walker-Engstrom's analysis was valid only in the UPPP subjects ( $n = 43$ , power = 50%), because this type of analysis can only be performed with a treatment that does not rely on device usage for effect.



**Fig 2.** Scatterplot of improvement in Functional Outcome of Sleep Questionnaire (FOSQ) and apnea-hypopnea index (AHI) after active or placebo treatment ( $n = 51$ ). Positive numbers denote improvement in AHI (ie, lower AHI after treatment) and FOSQ (ie, higher FOSQ score after treatment). Note the apparent lack of association between these variables.

While one can measure PSG parameters with a treatment device in place, it does not translate into the patient's day-to-day physiology unless the patient uses the device 100% of sleep time. It is unknown how to correct for incomplete device usage in estimating the actual physiologic effect of the device. Thus, there is no way to measure true PSG outcomes accurately for device patients.<sup>23</sup> With surgical therapy, however, the treatment effect does not rely on usage, so PSG outcomes can be measured. Thus, a surgical patient sample allows for this type of comparison of outcome mea-

tures. Similarly, untreated patients can be compared over time, but the sample must include subjects with some change in order to detect a correlation. A group treated with stable weight loss could be used to compare outcomes, but it is difficult to succeed with this mode of treatment.

This study is limited by the size of the sample of patients that had complete outcome measures ( $n = 54$ ), which provides only 60% power to detect an important correlation of 0.3. However, the consistently poor correlation among outcome measures suggests that the lack of consistent positive correlation is real. In addition, there was >80% power to rule out an important correlation in the baseline measures.

The sample of patients had mild to moderate OSAS, on average. Patients with more severe OSAS may have better correlation between PSG and non-PSG measures. However, prior research has shown a poor correlation between baseline measures in severe OSAS patients.<sup>7,8</sup> The relatively mild treatment effects in our sample may have dampened our ability to detect associations between PSG and non-PSG outcomes. However, subgroup analysis of only those receiving active treatment also revealed poor correlations (data not shown). Moreover, subgroup analyses of those with a "successful" PSG outcome (defined by either the combination of AHI decrease  $\geq 50\%$  and final AHI  $< 20$ ; or of apnea index decrease  $\geq 50\%$  and final apnea index  $< 10$ ) revealed poor correlations as well (data not shown). It would be useful to duplicate these analyses in a sample of severe OSAS patients who experience greater degrees of improvement in PSG and non-PSG outcomes.

Our regression analyses strengthen the findings of unimportant associations between PSG and non-PSG measures at baseline and at outcome. Adjustment for potential confounding variables (age, sex, and body mass index) help isolate the comparison of interest. However, we did not adjust for comorbidity, which may confound the association between PSG and non-PSG measures. Previous studies had similar findings for the comparison of baseline measures, even after adjusting for comorbidity. There may be other confounding variables as well.

## CONCLUSION

PSG indexes are not consistently associated with sleepiness, OSAS-related quality of life, general health status, or reaction times, either at baseline or as outcome measures in patients with mild to moderate OSAS. PSG indexes may not quantify some important aspects of OSAS disease burden or treatment outcome.

AHI in particular is a poor surrogate for sleepiness, OSAS-related quality of life, general health status, and reaction times. Clinically important outcomes should be measured directly.

## REFERENCES

1. Indications and standards for cardiopulmonary sleep studies. American Thoracic Society. Medical section of the American Lung Association. *Am Rev Respir Dis* 1989;139:559-68.
2. 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.
3. Peppard PE, Young T, Palta M, et al. Prospective study of the association between sleep-disordered breathing and hypertension. *N Engl J Med* 2000;342:1378-84.
4. Nieto FJ, Young TB, Lind BK, et al. Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study. Sleep Heart Health Study. *JAMA* 2000;283:1829-36.
5. Shepard JW Jr. Hypertension, cardiac arrhythmias, myocardial infarction, and stroke in relation to obstructive sleep apnea. *Clin Chest Med* 1992;13:437-58.
6. He J, Kryger MH, Zorick FJ, et al. Mortality and apnea index in obstructive sleep apnea: experience in 385 male patients. *Chest* 1988;94:9-14.
7. Weaver EM, Kapur V, Yueh B. Polysomnography vs self-reported measures in patients with sleep apnea. *Arch Otolaryngol Head Neck Surg* 2004;130:453-8.
8. Weaver EM, Patrick DL, Deyo RA, et al. Sleep study and quality of life correlation in sleep apnea (Abstract). *Otolaryngol Head Neck Surg* 2000;123:106.
9. Woodson BT, Steward DL, Weaver EM, et al. A randomized trial of temperature-controlled radiofrequency, continuous positive airway pressure, and placebo for obstructive sleep apnea syndrome. *Otolaryngol Head Neck Surg* 2003;128:848-61.
10. Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Los Angeles: Brain Information Service/Brain Research Institute, UCLA, 1968.
11. EEG arousals: rules and examples. a preliminary report from the disorders atlas task force of the American Sleep Disorders Association. *Sleep* 1992;15:174-84.
12. Weaver TE, Laizner AM, Evans LK, et al. An instrument to measure functional status outcomes for disorders of excessive sleepiness. *Sleep* 1997;20:835-43.
13. Piccirillo JF, Gates GA, White DL, et al. Obstructive sleep apnea treatment outcomes pilot study. *Otolaryngol Head Neck Surg* 1998;118:833-44.
14. Ware JE Jr, Snow KK, Kosinski M, et al. SF-36 health survey: manual and interpretation guide. Boston, MA: The Health Institute, New England Medical Center, 1997.
15. Johns MW. Daytime sleepiness, snoring, and obstructive sleep apnea: the Epworth Sleepiness Scale. *Chest* 1993;103:30-6.
16. Jewett ME, Dijk DJ, Kronauer RE, et al. Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. *Sleep* 1999;22:171-9.
17. Primer on correlation coefficients. *Effective Clinical Practice* 2001;4:139-40.
18. Hulley SB, Cummings SR. Designing clinical research: an epidemiologic approach. Baltimore, MD: Williams & Wilkins; 1988. pp. 247.
19. Stewart MG, Sicard MW, Piccirillo JF, et al. Severity staging in chronic sinusitis: are CT scan findings related to patient symptoms? *Am J Rhinol* 1999;13:161-7.
20. Brand PL, Duiverman EJ, Waalkens HJ, et al. Peak flow variation in childhood asthma: correlation with symptoms, airways

- obstruction, and hyperresponsiveness during long-term treatment with inhaled corticosteroids. Dutch CNSLD study group. *Thorax* 1999;54:103-7.
21. Wiklund I, Comerford MB, Dimenas E. The relationship between exercise tolerance and quality of life in angina pectoris. *Clin Cardiol* 1991;14:204-8.
  22. Walker-Engstrom ML, Wilhelmsson B, Tegelberg A, et al. Quality of life assessment of treatment with dental appliance or UPPP in patients with mild to moderate obstructive sleep apnoea: a prospective randomized 1-year follow-up study. *J Sleep Res* 2000;9:303-8.
  23. Weaver EM. Sleep apnea devices and sleep apnea surgery should be compared on effectiveness, not efficacy (letter). *Chest* 2003; 123:961-2.