CLINICAL EPIDEMIOLOGY/OUTCOMES RESEARCH

JAY F. PICCIRILLO, MD
Clinical Epidemiology/Outcomes Research Associate Editor

Obstructive sleep apnea treatment outcomes pilot study

JAY F. PICCIRILLO, MD, GEORGE A. GATES, MD, DONNA L. WHITE, BA, and KENNETH B. SCHECTMAN, PhD,
St. Louis, Missouri, and Seattle, Washington

Approximately 40 million Americans have chronic sleep disorders, the most serious of which is obstructive sleep apnea. The goals of this research were to serve as a demonstration project for a multicenter treatment outcomes research project for patients with obstructive sleep apnea. A clinical-severity staging system was created to control for important differences in the severity of sleep apnea among the enrolled patients. A disease-specific quality-of-life measure was used in this project to measure, from the patient's perspective, important pretreatment and posttreatment physical, functional, and emotional aspects of obstructive sleep apnea. Adults with apnea indexes greater than 5 who had not previously undergone uvulopalatoplasty were eligible. In total 142 patients were enrolled from eight otolaryngology practices. The mean age was 48 years, 112 were men, and 114 were white. The mean pretreatment apnea index was 40.0, and the mean respiratory distress index was 60.5. Seventy-one patients received continuous positive airway pressure, and 48 patients received surgery. Outcomes were assessed from scores on patient-based general and disease-specific health status measures 4 months after enrollment. The short duration of follow-up and limited number of patients undergoing posttreatment polysomnograms prohibit any analysis of treatment effectiveness. Nevertheless, this research represents a step forward for the support of future outcomes research projects by organized otolaryngology. (Otolaryngol Head Neck Surg 1998;118:833-44.)

According to the Executive Report of the National Commission on Sleep Disorders Research (NCSDR), approximately 40 million Americans have chronic sleep disorders, and another 20 million to 30 million have periodic sleep problems.[1] The most serious sleep disorder is obstructive sleep apnea (OSA) syndrome. The societal consequences of OSA are nontrivial. The NCSDR estimates that 75% of the 75,000 patients screened in accredited sleep laboratories each year will obtain a diagnosis of OSA, resulting in direct medical tests costing approximately $275 million per year.[1] The fragmentation of sleep caused by apneic events frequently contributes to excessive daytime sleepiness and has been associated with intellectual deterioration, behavioral and personality changes, enuresis, sexual dysfunction, and increased traffic accidents.[1-5] In its most severe form, OSA can lead to systemic or pulmonary hypertension,[6,7] cardiac arrhythmias, cor pulmonale, polycythemia, increased rate of occupational or driving accidents,[1] and increased mortality.[8] According to the NCSDR, 38,000 cardiovascular deaths as a result of OSA occur annually.

Many treatments currently are available for OSA, ranging from behavioral intervention (sleeping in a lateral position), diet and exercise programs, medications (e.g., stimulants or nasal steroid sprays), mandibular advancing dental devices (jaw repositioning or tongue retaining), continuous positive airway pressure (CPAP) or biventilation positive airway pressure, and a variety surgical procedures for removing or circumventing the

From the Department of Otolaryngology (Dr. Piccirillo and Ms. White) and Division of Biostatistics (Dr. Schectman), Washington University School of Medicine; and the Department of Otolaryngology–Head and Neck Surgery (Dr. Gates), University of Washington, Seattle.

Supported through a grant from the American Academy of Otolaryngology-Head and Neck Surgery Foundation, Inc.

Presented at the Annual Meeting of the American Academy of Otolaryngology–Head and Neck Surgery, San Diego, Calif., Sept. 18-21, 1994.

Reprint requests: Jay F. Piccirillo, MD, Clinical Outcomes Research Office, Department of Otolaryngology–Head and Neck Surgery, Washington University School of Medicine, Box 8115, 517 S. Euclid Ave., St. Louis, MO 63110.

0194-5998/98/$5.00 + 0 **23/1/84703**

site of obstruction.[9] The principal nonsurgical treatment is CPAP, which acts as pneumatic splint to passively prop the airway open.[2,9-11] Although CPAP is a proven, effective treatment for OSA,[3,12-19] patient compliance with this bulky apparatus is inconsistent because of noise from the machine, poor mask fit, nasal irritation, and difficulty in traveling.[1,2,20-22] The most commonly performed surgical procedure for OSA is the uvulopalatopharyngoplasty (UPPP),[23,24] which involves the removal of the uvula, inferior rim of the soft palate, and when present, the tonsils. Earlier studies, using 50% reduction in the preoperative apnea index (AI) as the definition of success, found that UPPP succeeds in 50% of patients with OSA.[25] Surprisingly, although a majority of patients report subjective improvement after UPPP,[26] often this is not associated with an improvement in objective measures (e.g., AI or sleep stage architecture).

Outcomes research is the scientific study of the outcomes of diverse therapies that are used for a particular disease, condition, or illness.[27] The need to determine the outcomes of medical care is fundamental to the successful and cost-effective practice of medicine.[28,29] The key features of outcomes research are the study of the effects of all major therapies on a condition, the expanded definition of outcome, and the central role of the patient in treatment selection.[30,31] The goals of this type of research are to document treatment effectiveness and create treatment guidelines.[32-34] Treatment effectiveness implies the value or strength of a treatment as is used in the community under usual situations. Outcomes research also uses a broadened description of patient outcome, which includes health status, quality of life (QOL), and satisfaction with care.

The main goals of the Obstructive Sleep Apnea Treatment Outcome Pilot (OSATOPS) project were (1) to serve as a demonstration multicenter treatment effectiveness and patient outcomes research project sponsored by the American Academy of Otolaryngology–-Head and Neck Surgery Foundation (AAO-HNSF); (2) to validate the outcomes assessment methods developed for this study, including the clinical severity index and OSA patient-oriented severity index (OSAPOSI), a disease-specific, health-related QOL (HRQOL) index; and (3) to acquire pilot data to support a large-scale, formal study to assess the effectiveness of common treatments for OSA.

METHODS

Description of Study Design

OSATOPS was a multiinstitutional, prospective, observational outcomes research study. The research protocol was developed by several of the authors (G. A. G. and J. F. P.); Dr. Maureen Hannley, Director of Research of the AAO-HNSF; and members of the Outcomes Research Subcommittee of the AAO-HNSF. The protocol was approved by the Sleep Disorders Committee of the AAO-HNS. The study design included use of repeat measures taken from the patient, sleeping partner, physician, and laboratory at three distinct times: initial visit, zero-time (the date of initial treatment), and outcome (4 months after zero-time). The questionnaires were developed and field tested at the OSATOPS Data Coordinating Center at the Department of Otolaryngology, Washington University School of Medicine. Initial treatment was selected by patients based on physician recommendations.

Description of Study Sites

The 10 participating medical centers represented the broad spectrum of clinical otolaryngology practices (4 academic, 5 private practice, and 1 military) across the entire country (5 Midwest, 2 Northeast, 2 Southwest, 1 West Coast). All participating physicians were board-certified otolaryngologists and members of the AAO-HNS. The study protocol was approved by the Institutional Review Boards at each center before project initiation. The centers enrolled patients between March 1, 1993, and December 31, 1993.

Description of study population. Eligible patients were between 18 and 75 years old and fulfilled the following criterion for the clinical definition of OSA: all-night polysomnogram (PSG) confirmation of at least five episodes of apnea per hour (respiratory cessation lasting at least 10 seconds). Patients with either new or previous diagnoses were eligible. Patients were excluded from study participation if they had previous UPPP, neuromuscular disorders, previously diagnosed craniofacial syndrome, nasopharyngeal stenosis, maxillofacial trauma, acute illness, or impaired ability to give consent.

Description of Study Variables

Demographic. Standard demographic data included age, sex, race, highest educational level achieved, and employment status. Patients who indicated that they were disabled were asked whether this was because of their sleep disorder.

History of present illness. Patient-based questionnaires collected information on the type, duration, and severity of symptoms as well as the response to any previous treatment. Information was also obtained about the reason the patient sought medical care and about the patient's most bothersome symptom. Information on risk factors (smoking, alcohol use, and weight gain) was also obtained.

The degree of hypersomnolence was evaluated with the Epworth Sleepiness Scale (ESS).[35] The ESS is a valid patient-based instrument that quantifies the degree of hypersomnolence by rating the likelihood of dozing in eight different situations (sitting and reading, watching television, sitting inactive in a public place, riding as a passenger in a car for an hour without break, lying down to rest in the afternoon when circumstances permit, sitting and talking to someone, sitting quietly after a lunch without alcohol, and sitting in a car while stopped for a few minutes in traffic) on a four-category scale of 0 (never) to 3 (high chance of dozing).

General health status. General health status was evaluated by use of the Medical Outcomes Study Short-Form 36 (MOS SF-36).[36,37] The MOS SF-36 rates a patient's health and functional well-being in eight different domains: physical functioning, role disability because of physical problems (role-physical), bodily pain, general health, vitality, social functioning, role disability because of emotional problems (role-emotional), and mental health. Scores for each domain range from 0 to 100, with 0 being the worst and 100 being the best. Normative data for the SF-36 has been obtained through extensive research and wide general usage.[38] Responses to the SF-36 were scored according to the algorithm as recommended by Hays et al.[39]

Disease-specific HRQOL. A measure the disease-specific HRQOL was obtained by the score on the OSAPOSI. We created the OSAPOSI especially for this study by first conducting multiple semistructured interviews with patients to identify the cogent physical problems, functional limitations, and emotional consequences of OSA and its treatment. The OSAPOSI includes the 32 items that were most frequently mentioned and seemed clinically meaningful (Table 1). The items (e.g., difficulty staying awake) were organized into five problem subscales (sleep, awake, medical, emotional and personal, and occupational). The OSAPOSI has two categoric scales so patients can rate the magnitude of the problem for each item (magnitude scale: 0 to 5, from no problem to problem is as "bad as can be") and the importance to the patient (importance scale: 1 to 4, from not important to extremely important). A symptom-impact score is calculated as the product of the magnitude score and the importance score; the higher the score is, the worse the HRQOL. The range of the score for any one item is 0 to 20, and for the entire instrument the range is 0 to 640. Finally, patients are asked to provide a global rating of the overall amount of "bother" or "disturbance" they experience as a result of OSA.

The OSAPOSI provides different information from the PSG and other laboratory-based measures of OSA.

Table 1. Items on OSAPOSI

Sleep problems	
1.	Trouble falling asleep
2.	Waking during sleep
3.	Loud/excessive snoring
4.	Restlessness during sleep
5.	Waking "too early" in morning
6.	Waking up feeling tired
7.	Bed wetting
Awake problems	
8.	Fatigue or tiredness
9.	Frequent yawning
10.	Sleepiness while driving
11.	Memory and/or concentration problems
12.	Productivity limited at certain times of day
13.	Often late for meetings or appointments
14.	Participation in community, volunteer, religious, or spiritual activities limited
Medical problems	
15.	Amount of medical care required for OSA
16.	Interaction of OSA with other medical problems
17.	Travel by automobile to other regions or parts of country limited because of fear of medical problem
18.	Unable to have sexual relations because of medical problem
19.	Financial burden as a result of illness
Emotional and personal problems	
20.	Dread/fear going to bed
21.	Nerves are "right on surface"
22.	Inability to relax, always anxious
23.	Marital strain, stress, and tension
24.	"Foul" mood
25.	Unable to experience closeness with spouse and/or others
26.	Lack of desire for sexual relations
27.	Feeling that future is hopeless
Occupational impact	
28.	Competence questioned
29.	Reliability questioned
30.	Inability or difficulty getting new job
31.	Loss of job
32.	Modification in job because of excessive sleepiness

Despite the fact that patients may underreport or misreport their symptoms, patient-based measures, like the OSAPOSI, provide important information with a different perspective on the patient's condition. Preliminary research indicated that the OSAPOSI offers promise as a valid and sensitive patient-based assessment of the QOL for patients with OSA.[40]

Physician. The physician recorded data in four main areas: vital signs, physical examination, Müller maneuver,[41] and medical comorbidities. During physical examination the physician rated the amount of dentition, presence of a draping soft palate, elongated uvula, redundant pharyngeal folds, and abnormalities of the epiglottis and hypopharynx. The degree of turbinate swelling, presence and severity of septal abnormalities, Mallampati classification[42] of the oropharynx, and quality of the false vocal cords were also recorded. The Mallampati classification rates the severity of oropha-

A

Redundant pharyngeal tissue	BMI <30	BMI 30-40	BMI >40
Absent	Alpha	Beta	Gamma
Present	Alpha	Gamma	Gamma

B

ESS	Physical-severity index: Alpha	Physical-severity index: Beta	Physical-severity index: Gamma
<9	A	A	B
9-16	A	B	C
>16	B	C	C

C

Minimum O_2 saturation (%)	RDI 0-33	RDI 34-65	RDI >65
>84	1	2	2
84-65	1	2	3
<65	2	2	3

D

Functional-severity index	PSG-severity index: 1	PSG-severity index: 2	PSG-severity index: 3
A	I	I	III
B	II	II	III
C	II	III	III

Fig. 1. Creation of clinical-severity staging system. Panels **A** through **C** demonstrate the sequential conjunction and consolidation of key physical examination variables, ESS, and PSG variables to ultimately create the clinical-severity index. **A,** Pattern of consolidation of redundant pharyngeal tissue and BMI to form composite physical-severity index. Categories of BMI and redundant pharyngeal tissue are conjoined to create the three-category (alpha, beta, and gamma) physical-severity index. **B,** Pattern of consolidation of ESS and physical-severity index to form composite functional-severity index. Categories of physical-severity index (alpha, beta, and gamma) are conjoined with three categories of the ESS (<9, 9 to 16, >16) to create the functional-severity index. **C,** Pattern of consolidation of minimum O_2 saturation during apnea and RDI to form the composite PSG-severity index. Categories of the two key PSG variables, minimum O_2 saturation during apnea and RDI, are conjoined to create the three-category (1, 2, and 3) PSG-severity index. **D,** Pattern of consolidation of functional-severity index and PSG-severity index to form the composite clinical-severity index. The three categories (A, B, and C) of the functional-severity index and three categories (1, 2, and 3) of the PSG-severity index are conjoined to create the three-category (I, II, and III) composite clinical-severity index.

ryngeal obstruction on a three-category pictographic scale.

The Müller maneuver was performed with patients in both the seated and supine positions. The site of obstruction was classified as the velopharynx only if the velopharynx or tonsil regions were rated as 50% or more collapsed and the epiglottis and base of the tongue were rated as less than 50% collapsed. The site of obstruction was classified as the hypopharynx only if the epiglottis or base-of-tongue region was 50% or more collapsed and the tonsil and velopharynx were less than 50% collapsed. If the velopharynx/tonsil and epiglottis/base-of-tongue regions were both rated as 50% or more collapsed, the site of obstruction was classified as being both the velopharynx and hypopharynx.

The physician listed any other diseases or conditions the patients had and rated the severity using the Kaplan-Feinstein comorbidity index.[43] The CAGE questionnaire[44,45] was used to determine the likelihood of alcoholism. The physician also rated the overall impression of the severity of the OSA on a five-category rating scale (0 to 4, none to severe).

Laboratory. All-night PSG and a next-day Multiple Sleep Latency Test scores were required for all patients. The PSG and Multiple Sleep Latency Test data were recorded and scored by the professionals associated with the sleep laboratory connected with each participating center using definitional criteria approved by the American Sleep Disorders Association.[46,47]

Creation of clinical-severity staging system. To assess treatment effectiveness from an observational study, controls for important prognostic and therapeutic differences among patients must be used.[48] At present, no validated prognostic staging system for OSA severity exists. To develop such a system, we hypothesized that an ideal prognostic staging system would likely contain descriptions of the degree of daytime sleepiness, severity of physical examination abnormalities, and derangements on various physiologic parameters obtained from the all-night PSG. Furthermore, it was believed that the severity of illness should agree with the physician's assessment of severity, rather than the patient's, because the physician's severity estimates likely reflect more pathophysiologic severity. The dependent variable (or gold standard) used to create the index was the physician's overall impression of the

Table 2. Mean SF-36 and OSA patient-oriented severity scales as a function of clinical-severity stage

Clinical severity stage	Physical functioning*	Role-physical†	Pain‡	General health‡	Vitality	Social functioning	Mental health	Role-emotional†	OSAPOSI
I	91.8	83.3	86.5	72.7	44.0	56.3	68.7	95.8	3.25
II	77.3	61.1	77.4	60.1	40.1	52.2	69.2	72.2	3.72
III	66.2	47.4	69.4	57.1	35.1	50.7	67.9	61.4	4.77
TOTAL	73.0	56.4	74.0	60.1	37.8	51.9	68.4	69.4	4.26

*$p < 0.001$.
†$p < 0.01$.
‡$p < 0.05$.

severity of the OSA. The analytic plan was to identify those baseline subjective, physical examination, and PSG variables that were related to the physician's assessment of severity and then to combine these variables into a multivariate staging system. The development of the clinical-severity staging system will be discussed briefly.

The five independent variables that were included in the index were the ESS, body mass index (BMI), presence of redundant pharyngeal tissue, respiratory distress index (RDI), and minimum O_2 saturation during apnea. These five significant variables were merged in a series of steps by use of the principles of conjunctive consolidation[49] to create a three-stage composite clinical-severity index. The various combinations of the five significant variables that define each clinical-severity stage are demonstrated in Fig. l. The following example demonstrates how a particular stage is determined for a patient. For instance, a patient with an ESS score of 10, no redundant pharyngeal tissue, a BMI of 35, a minimum O_2 saturation level of 85, and an RDI of 40 would be classified as having a clinical-severity stage of II.

To assess the concurrent validity of this newly created index, we compared mean scores on each of the eight domains of the SF-36 within each clinical-severity stage. As shown Table 2, the mean SF-36 scores for physical functioning, role-physical, bodily pain, and general health were significantly different across the three clinical-severity stages.

Treatment. The patient provided treatment information for two time intervals: previous (treatment received before initial visit) and current (treatment received at the time of initial visit). The physician recorded the type of initial treatment given to the patient. If surgery was performed during participation in this research project, data were obtained on the type of procedure, duration, blood loss, and development of perioperative and postoperative complications, if any.

Outcome. Four months after treatment, patients completed the MOS SF-36 and the OSAPOSI, indicated why they chose a particular treatment and whether they would choose that same treatment again, and gave information on the success of their treatment. The patients were also asked whether they had any complications from treatment and, if so, how severe they were. Finally, patients who received CPAP rated their compliance by approximating both the number of nights they went to bed with CPAP and the number of mornings they awoke with the CPAP mask in place.

Physicians recorded their perceptions of the success and, when appropriate, patient compliance with treatment. Physicians also recorded any complications the patients had.

Information from the spouse or bed partner about the patient's snoring, sleep apnea, overall degree of bother, and whether the spouse or bed partner was able to sleep in the same bed was also obtained.

Statistical and analytic plan. Frequency counts were used for baseline description of the study centers and the population. Bivariate statistics were performed with Student's *t* test, correlation statistics,[50] the χ^2 test, and analysis of variance. Cronbach's alpha coefficient[51] was used to measure construct validity of the OSAPOSI by evaluation of the degree of interitem correlation on the OSAPOSI. Multivariate analyses, including multiple logistic regression and conjunctive consolidation, were used to construct the clinical-severity staging system and to assess the prognostic value of selected variables. Conjunctive consolidation is an alternative to regression analysis for the study of multiple variables.[49] Conjunctive consolidation often performs as well as logistic regression for predictive modeling and has the advantage of being easier to implement in a clinical setting.[52] All tests were two tailed, and statistical significance was established at the $p < 0.05$ level.

RESULTS

Study Centers

There were 10 study centers in this project. Four centers enrolled more than 80% of the total enrolled population; two centers failed to enroll a patient.

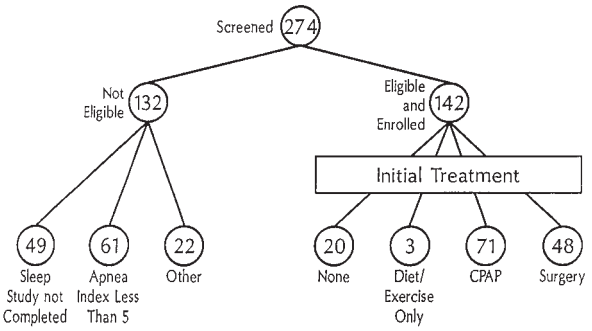

Fig. 2. Flow diagram of patient participation. Numbers in circles represent number of patients.

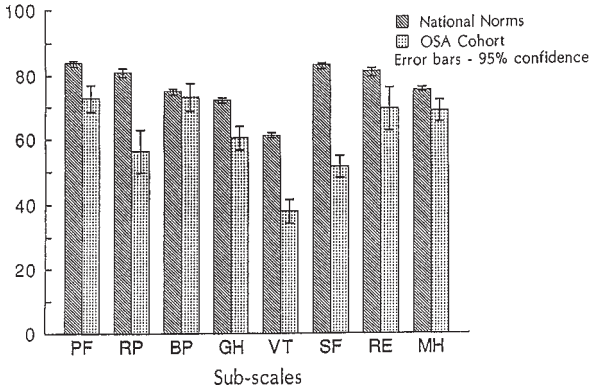

Fig. 3. General health status and MOS SF-36 scores for national norms and OSA cohort. Subscales: *PF,* physical functioning; *RP,* role-physical; *BP,* bodily pain; *GH,* general health; *VT,* vitality; *SF,* social functioning; *RE,* role-emotional; *MH,* mental health. Scores on each subscale range from 0 to 100; higher scores indicate better function.

Table 3. Baseline description of population

Variables	No. of patients	%
Age		
<41 yr	38	27
41-57 yr	72	51
>57 yr	32	22
Gender		
Male	126	89
Female	16	11
Race		
Black	18	13
White	114	80
Other	10	7
Education level		
College graduate	51	36
High school graduate	69	48
<12 yr	17	12
Unknown	5	4
Employment status*		
Working full-time	98	—
Working part-time	9	—
Keeping house	3	—
Retired	18	—
Student	6	—
Disabled	19	—
Disabled by apnea	6	—

*Numbers total more than 142 because patients could select more than one category.

Study Population

Baseline description. The description of the population is shown in Table 3. In total, 274 patients were screened, of whom 142 were eventually enrolled. The flow diagram of patient participation is shown in Fig. 2. The mean age ± SD of the enrolled population was 48 ± 12.1 years; 114 (80%) patients were white, and 126 (89%) were men. More than a third of the subjects were college graduates. Most were working full-time or part-time. Nineteen patients indicated that they were disabled, 6 because of their sleep apnea.

History of present illness. Most patients reported fatigue as their chief reason for seeking medical attention. Other reasons included breathing problems during sleep and snoring. The average duration of symptoms before the initial visit was 11 months (range, 0.5 to 40 months). Most patients' (77%) OSA was newly diagnosed, and they had not yet received treatment. The mean ESS[35] response was 13.8, the range was 1 to 24, and the interquartile range was 8. Of note, the correlation (Pearson's correlation coefficient, r) between the ESS and the RDI was 0.20 ($p = 0.02$).

General health status. The SF-36 scores for the patients with OSA and a cohort of adult patients comprising national norms[38] are displayed in Fig. 3. When compared with national norms, all subscales except bodily pain were significantly worse in the patients with OSA.

Disease-specific HRQOL. The pretreatment disease-specific HRQOL symptom-impact scores from the OSAPOSI are shown in Fig. 4. The mean ± SD pretreatment total instrument score was 4.2 ± 2.8; the range was 0.03 (best) to 17.1 (worst). The sleep and awake subscales had the highest mean scores (6.7 and 6.3, respectively), indicating that these domains were most affected by OSA.

Physical examination, comorbidity, and overall severity. The mean ± SD BMI was 34.7 ± 9.0; 28 (20%) patients had BMIs greater than 40 (i.e., morbidly obese). One hundred eleven (79%) patients had draping soft palates, 110 (77%) had elongated uvulae, and 99 (70%) had redundancies of the pharyngeal tissue. On the basis of the Mallampati classifications, 4 (3%)

patients had a normal pharynxes, 58 (42%) had moderately constricted pharynxes, and 76 (55%) had severely constricted pharynxes. The Müller maneuver was performed in 129 patients. Seventy (54%) patients were classified as having only velopharyngeal sites of obstruction, 58 (45%) as having both velopharyngeal and hypopharyngeal sites of obstruction, and 1 (1%) as having no site of obstruction. The systolic blood pressure was above 140 mm Hg in 32 (22%) patients, and the diastolic blood pressure was above 90 mm Hg in 28 (20%) patients.

The most frequently reported medical comorbidity was hypertension. The ratings of medical comorbidities according to the Kaplan-Feinstein index were as follows: 57 (45%) patients had no comorbidities, and 58 (46%), 9 (7%), and 3 (2%) had mild, moderate, and severe comorbidities, respectively. On the basis of the responses to the CAGE questionnaire,[44,45] 12 (8%) patients were likely to be alcoholics.

Physicians rated the severity of OSA in 9 (7%) as very mild, 40 (31%) as mild, 49 (38%) as moderate, and 31 (24%) as severe. The mean ± SD AI and RDI were 38.0 ± 32.1 and 60.5 ± 36.2, respectively.

Validation of the OSAPOSI. The validity of the OSAPOSI was assessed in several ways. First, concurrent validity was assessed by comparison of scores on the OSAPOSI at baseline with the patient's response on the overall global rating of disease-specific QOL. The mean total symptom-impact score was significantly related to the overall global rating (F value = 17.39; $p < 0.0001$). Next, construct validity was measured with Cronbach's alpha analysis.[51] The Cronbach's alpha score was 0.93, indicating excellent interitem correlation.

Responsiveness to change was measured by comparison of the change (pretreatment from posttreatment) in the OSAPOSI score with the patients' global responses to treatment at 4 months. The changes in the OSAPOSI scores correlated with the patients' overall assessments of their responses to treatment. This relationship was statistically significant (F value = 4.95; $p < 0.001$).

Sensitivity to change was assessed with the standardized response mean (SRM),[53] and the magnitude of change was assessed with the effect size (ES).[54] The SRM, calculated by dividing the mean change score by the standard deviation of the change, was 0.63. The higher the SRM value, the greater the sensitivity to change. The ES, calculated as the mean change divided by the standard deviation of the baseline score, was 0.59. The higher the ES, the greater the magnitude of change. These scores suggest that the OSAPOSI is sensitive to clinical change.

Next, a clinically meaningful difference in OSAPOSI scores was calculated by analysis of the mean OSAPOSI difference scores within categories of response to treatment (i.e., much improved, somewhat improved, no change, somewhat worse, much worse). As a result of research not shown here, it was determined that a change in OSAPOSI score of 50% was clinically meaningful. That is, it was very likely that a patient would indicate that he or she was much improved or somewhat improved on the global rating of treatment effectiveness if his or her posttreatment OSAPOSI score decreased 50% from the pretreatment value. On the basis of this definition, 42 of the 83 evaluable patients achieved significant improvement in their disease-specific QOL ratings.

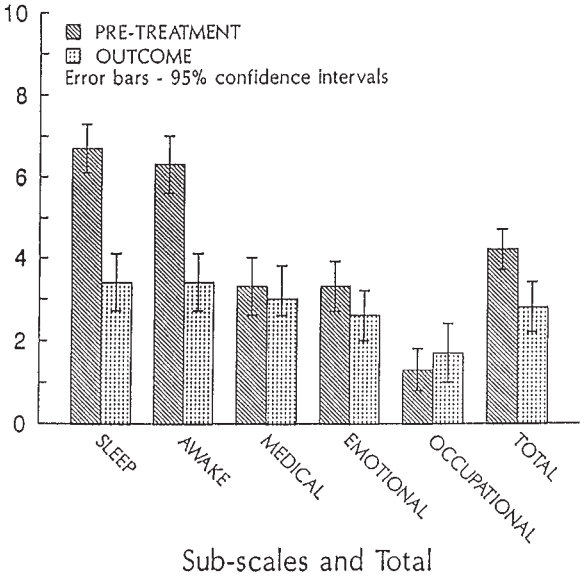

Fig. 4. Disease-specific health status and HRQOL. OSAPOSI pretreatment and outcome scores.

Initial treatment. Of the 142 eligible and enrolled patients, 122 (85.9%) received some form of initial treatment: 71 (50.0%) received CPAP, 48 (33.8%) surgery, and 3 (2.1%) diet/exercise (Fig. 2). Of the 48 patients who elected surgery, 44 (92%) received UPPP, 1 (2%) septoplasty, 1 (2%) tonsillectomy, 1 (2%) genioglossus advancement, and 1 (2%) laser-assisted uvulopalatoplasty. Of the 44 patients who received UPPP, 15 (34%) also received concomitant septoplasties.

Physicians were asked to indicate to what extent (largely, partially, minimally, or not at all) treatment recommendations were based on the information obtained from history, physical examination, and laboratory findings. Physicians were also asked to indicate to what extent treatment recommendations were based on patient preferences and insurance status. Patient preference was listed as the most important reason for

Table 4. SF-36 scores

Category	Pretreatment		Posttreatment		Difference	
	Mean	SD	Mean	SD	Mean	SD
Physical functioning	73.7	26.9	75.6	25.4	2.5	23.3
Role-physical*	55.7	40.6	70.2	38.0	15.4	42.2
Pain	72.8	25.4	72.5	25.7	–0.35	18.4
General health	61.4	20.9	60.6	23.4	–0.37	17.0
Energy/fatigue*	39.8	22.5	57.8	22.1	17.9	27.0
Social functioning	51.9	20.5	53.5	17.5	1.28	22.9
Role-emotional	69.2	40.0	74.9	37.9	5.38	32.2
Emotional well-being*	68.4	21.2	73.5	18.9	5.19	13.5

*$p < 0.001$.

the treatment decision for 49% of the patients, whereas results from the sleep study were the most important reason for 22%. Only one patient had treatment recommendations based largely on insurance reasons.

Follow-up. Of the 122 patients who received some form of treatment, 77 (63%) underwent repeat PSG, and 96 (79%) completed questionnaires at 4 months. Of the 71 patients receiving CPAP, 28 (39%) underwent follow-up PSGs, and 54 (76%) returned the questionnaires. Of the 48 patients who underwent surgery, 21 (44%) underwent postoperative PSGs, and 31 (65%) returned the questionnaires. The reasons for the low rate of repeat PSG included inability to obtain insurance coverage for the cost of the study and patient unwillingness to return for overnight study. There was no difference in age, sex, BMI, ESS, site of collapse, physician's rating of severity, and clinical-severity stage between those patients who did and those patients who did not have follow-up PSGs.

Outcome. Because of the low rate of follow-up PSG, outcomes results are reported on the basis of the questionnaire responses only.

SF-36 scores. The difference in mean SF-36 individual domain scores between baseline and 4 months was compared with the paired Student's *t* test statistic, and the results are displayed in Table 4. Scores on the role-physical, energy/fatigue, and emotional well-being subscales of the SF-36 increased most significantly from baseline.

OSAPOSI scores. Figure 4 shows the pretreatment and outcome scores for the OSAPOSI. As can be seen, scores on the sleep and awake subscales and the total instrument score changed the most.

Complications of therapy were reported by 32 patients; the rates were not significantly different between the CPAP (18 of 50, 36%) and the surgical groups (14 of 30, 47%; $\chi^2 = 0.889$, $p = 0.346$). Of the 32 patients who reported side effects, 9 rated them as extremely bad or very bad. The 7 patients receiving CPAP who rated their side effects as being this severe listed dizziness, teeth grinding, discomfort/irritation, and rhinorrhea as their side effects, whereas the 2 patients who had undergone surgery listed regurgitation, voice change, numbness, swallowing difficulty, breathing difficulty, infection, swelling, and sore throat. Physician reports of anesthesia- and surgery-related complications were few. Intubation was reported as being difficult in 2 patients. Postoperative airway compromise developed in 2 patients; 1 required an emergent tracheotomy, and the other was reintubated successfully. Other complications included bleeding (2 patients), pneumothorax (1 patient), and postoperative edema (1 patient).

Evaluation of treatment effectiveness. The third aim of this project was to acquire pilot data to assess treatment effectiveness. The protocol stipulated that each patient would have treatment effects measured objectively with PSG and subjectively with the MOS SF-36 and OSAPOSI. Unfortunately, as described previously, only a minority of patients underwent repeat PSG, and therefore the results from this test could not be used as an outcome measure. Because of the lack of an objective measure and the small number of patients in the different treatment groups, the investigators believed it was inappropriate to evaluate treatment effectiveness.

Practice variation across study centers. Because this was a multiinstitutional study including a variety of different practice settings from across the United States, analyses of variations in treatments and outcomes were performed. The analyses were restricted to those centers that enrolled more than seven patients each. These centers accounted for 70% of the total patients enrolled. The results are shown in Table 5.

Types of patients. The overall percentage of patients whose conditions were newly diagnosed (ver-

Table 5. Baseline, treatment, and outcome variables across study centers

Study center	% New cases (*n*)*	BMI (mean ± SD)†	RDI (mean ± SD)†	% Clinical severity stage 3 (*n*)*	% Surgical cases (*n*)	OSAPOSI difference score (mean ± SD)	% Improvement rate (*n*/total)
A	100 (9)	28.4 ± 3.1	27.1 ± 14.2)	0 (—)	70 (7)	1.3 ± 2.6	17 (1/6)
B	81 (13)	38.2 ± 7.8	67.5 ± 48.8	61 (11)	33 (6)	2.1 ± 2.2	53 (8/15)
C	69 (18)	33.8 ± 8.0	55.8 ± 36.5	50 (19)	12 (5)	1.2 ± 5	42 (13/31)
D	44 (10)	38.7 ± 15.0	72.8 ± 38.6	76 (13)	33 (8)	0.86 ± 2.5	64 (9/14)
E	94 (30)	34.9 ± 6.7	61.5 ± 23.5	64 (14)	39 (14)	2.6 ± 2.2	56 (15/27)
TOTAL	75 (80)	35.2 ± 9.3	59.5 ± 36.1	58 (57)	31 (40)	1.7 ± 2.7	49 (46/93)

*$p < 0.001$.
†$p < 0.05$.

sus those whose conditions had been previously diagnosed) who were enrolled was 75%, and the rate varied from 44% to 100%. The mean BMI was 35.2, with values ranging from a low of 28.4 to a high of 38.7. The mean RDI ranged from 27.1 to 72.8. The percentage of patients whose OSA was classified as clinical-severity stage 3 (worst group) ranged from 0% to 76%. These differences in the types of patients enrolled across study centers were all statistically significant.

Treatment selection. Of the 142 patients enrolled, 71 (50%) received CPAP and 48 (34%) surgery. The rate of UPPP varied between 12% and 70% across study centers. To further investigate this wide difference, the association between baseline patient and physician factors and initial treatment was examined. Predictors of initial treatment choice (CPAP or surgery) in this population included age group ($\chi^2 = 9.142$, $p < 0.010$), category of obesity as determined by BMI ($\chi^2 = 8.005$, $p < 0.018$), and physician ($\chi^2 = 28.082$, $p = < 0.002$). Because multiple variables were associated with initial treatment choice, multivariate analysis was performed to identify the most important treatment predictors. Age group and degree of obesity were both important predictors of initial treatment; the older and heavier patients were more likely to be treated with CPAP. After the age and weight of the patient were controlled for, treatment selection was no longer related to the physician. Among the six otolaryngologists who performed the most operations, the rate of combination of UPPP and septoplasty varied from 0% (0 of 5) to 83% (5 of 6). This difference in the rate of the use of septoplasty with UPPP across study centers was statistically significant ($\chi^2 = 15.8$, $p = 0.008$). Interestingly, among the 44 patients who underwent UPPP, 33 had septal deviations. Of these 33 patients, 16 underwent concomitant septoplasty, and the other 17 underwent UPPP alone. Therefore we believe this difference in the rate of the use of septoplasty with UPPP across study centers, without a concomitant difference in prevalence and severity of septal deformities, represents true practice variation.

DISCUSSION

This AAO-HNSF–supported project represents the first step forward in the conduct of outcomes research by organized otolaryngology. The AAO-HNSF provided direct scientific support for the development of the project protocol and monetary support for various aspects of the project, including data collection and analysis. The OSATOPS demonstrates the ability of the AAO-HNSF to support multiinstitutional, observational outcomes research.

The second aim of this project was the validation of outcomes assessment methods and patient-based measures of outcome. The clinical-severity staging system provides a rational method of risk adjustment or prognostic stratification, which is critically important in observational research. Data collected in this project allowed for the creation of such a system. Although validation of the clinical-severity staging system awaits future research efforts, the development of this system must be considered an achievement of this project. The OSAPOSI is a disease-specific HRQOL index created especially for this project. It appears that the OSAPOSI is a valid health status and QOL measure and should be incorporated into future research projects. Additional clinimetric research (not shown here) demonstrated that a difference of 0.50 on the OSAPOSI was equivalent to a clinically meaningful difference. On the basis of this information and the observed differences in OSAPOSI scores and standard deviations between the CPAP and UPPP groups, sample size calculations suggest that nearly 3000 patients per treatment group would need to be enrolled for future studies of treatment effectiveness. Unfortunately, the large number of patients makes the conduct of a traditional clinical trial extremely expensive and time consuming. Alternative designs, such as

an observational study, are likely to be more successful at answering the important clinical questions.

The third aim of this project was the evaluation of treatment effectiveness. As a result of the lack of randomization, short duration of follow-up, and the fact that less than 50% of patients underwent postoperative PSGs, definitive conclusions about treatment effectiveness are impossible. We believe that it is imperative that future studies have sufficient funds to provide remuneration to patients for both the cost of the PSGs and the time required for these studies.

There was a significant amount of variation in the types of patients included in this study and the use of surgical treatments for OSA across treatment centers. This heterogeneity underscores the importance of adequate risk and severity adjustment methods before the interpretation of patient outcomes and evaluation of treatment effectiveness from observational research. The practice variation in the use of procedures is consistent with that of previous studies of other, nonotolaryngologic, conditions.[55,56] This finding is not particularly surprising given the already recognized variations in practice style across different practice settings and geographic regions.[27,55-57] The degree and impact of this variation on outcome, quality, and cost is unknown at this time and should be the focus of future investigations. To improve the scientific conduct of future studies, the performance of surgical procedures should be standardized across study centers, and compliance monitors should be used with CPAP machines.

The limitations of this study must be considered when the results are interpreted. First, the OSATOPS was a 4-month pilot project and was not designed to provide definitive answers about treatment effectiveness. Second, patient follow-up was incomplete and possibly biased. Third, despite the use of standard criteria for the reporting of PSG information and the provision of explicit definitions of physical examination variables, the reporting of such information from different centers seemed, at times, inconsistent and nonstandardized. Sleep medicine is a relatively new field in which the standard terminology suggested for recording and then reporting data is not consistently used by all sleep laboratories. Further, some sleep laboratories strictly use computerized scoring algorithms, whereas others use partial or total technician "hand scoring." Although a standardized form was created by use of definitions recommended by the American Sleep Disorders Association for the reporting of PSGs, several centers provided information that was not consistent with standard definitions. We believe that in future multiinstitutional studies it will be critical that personnel agree to standard criteria for reading and recording. Ideally, all PSGs should be read at a designated "reading center" where interobserver variability can be minimized through aggressive quality-control efforts. Fifth, because of the excessive number of forms and questionnaires, patient and physician participation was made difficult, and data entry and analysis were protracted. This study has identified many variables, originally thought to be important, which were not found to be valuable in the preliminary analyses conducted in this study. A more streamlined approach to data collection is possible. Sixth, because of financial restrictions, standard policies for the conduct of multiinstitutional studies, including monitoring of data integrity and standardized reporting of information, were not implemented. The inability to provide patient remuneration certainly led to the high percentage of patients not receiving posttreatment PSGs. Although we did receive payment for each patient successfully enrolled, this amount did not cover the full cost of study center participation. Despite all these limitations, we still consider this pilot project to be important because it represents the first step toward future outcomes research in OSA. The design and conduct of subsequent studies can be more efficient as a result of this study.

We express profound appreciation to the many people whose diligent work made the successful completion of this project possible. In particular, we acknowledge the following individuals at Washington University who participated in this project: Dr. John Miller, Director of the Sleep Disorders Center, and Dr. Stephen Duntley, Department of Neurology, for developing of the sleep study data collection protocol; Dr. Allen Sclaroff, Department of Otolaryngology, and Dr. Franz Wippold, Mallinckrodt Institute of Radiology, for developing the cephalometric and CT topogram protocol and scoring all of the lateral cephalometric x rays and CT topograms; Dorothy Edwards, Program in Occupational Therapy, for assisting with the psychometric evaluation of the OSAPOSI; Nancy Kollmar, Clinical Outcomes Research Office, Department of Otolaryngology, for designing and managing the database; and Jack Baty, Division of Biostatistics, for managing and supporting the SAC data set.

The following AAO-HNS physicians acted as Study Center principal investigators: Tucker Woodson (Medical College of Wisconsin–Milwaukee), Jay Piccirillo (Washington University), Jonas Johnson (University of Pittsburgh), Victor Ejercito (Marshfield Clinic), Charles Johnson (Lackland AFB), William Cast (Fort Wayne), George Katsantonis (Park Central Institute), Martin Hopp (Cedars-

Sinai), David Eisele (Johns Hopkins), and William Moran (Oklahoma City).

Finally, many thanks for the tireless work of the research assistants at each study center: Sharon Gaynor, RN; Carol Beck, Research Coordinator; Jaya Reddy, Research Technician; Kim Kochman, Research Coordinator; Nita Whiteside, Research Coordinator; Sue Finley, RN; Helen Brown, LPN; and Olga Garza, Secretary. We also thank Janet Kelly (University of Washington) for help with data entry.

REFERENCES

1. National Commission on Sleep Disorders Research. Wake up America: a national sleep alert. Washington (DC): Government Printing Office; 1993.
2. Engleman HM, Martin SE, Deary I, Douglas NJ. Effect of continuous positive airway pressure treatment on daytime function in sleep apnea/hypopnea syndrome. Lancet 1994;343:572-4.
3. Rajagopal KR, Bennett LL, Dillard TA, Tellis CJ, Tenholder MF. Overnight nasal CPAP improves hypersomnolence in sleep apnea. Chest 1986;90:172-6.
4. Kribbs NB, Getsy JE, Dinges DF. Investigation and management of daytime sleepiness in sleep apnea. In: Saunders NA, Sullivan CE, editors. Sleep and breathing. 2nd ed. New York: Marcel Dekker; 1994. p. 575-604. Part of: Lenfant C, editor. Lung biology in health and disease. Vol 71. New York: Marcel Dekker; 1994.
5. George CF, Nickerson PW, Hanly PJ, Millar TW, Krygen MH. Sleep apnea patients have more accidents. Lancet 1987;2:227.
6. Fletcher EC, DeBehnke RD, Lovoi MS, Gorin B. Undiagnosed sleep apnea among patients with essential hypertension. Ann Intern Med 1985;105:190-4.
7. Kales A, Bixler EO, Cardieux RJ, et al. Sleep apnea in a hypertensive population. Lancet 1984;2:1005-8.
8. He J, Kryer MH, Zorick FJ, et al. Mortality and apnea index in obstructive sleep apnea. Experience in 385 male patients. Chest l988;94:9-14.
9. Westbrook PR. Sleep disorders and upper airway obstruction in adults. Otolaryngol Clin North Am 1990;23:723-43.
10. Sullivan CE, Grunstein RR. Continuous positive airway pressure in sleep-disordered breathing. In: Kryger MH, Roth T, Dement WC, editors. Principles and practices of sleep medicine. Philadelphia: WB Saunders; 1989. p. 559-70.
11. Strohl KP, Redline S. Nasal CPAP therapy, upper airway muscle activation and obstructive sleep apnea. Am Rev Respir Dis 1986;134:555-8.
12. Sullivan CE, Issa FG, Berthon-Jones M, McCauley VB, Costos LJV. Home treatment of obstructive sleep apnea with continuous positive airway pressure applied through a nose mask. Bulletin Européen de Physiopathologie Respiratoire 1984;20:49-54.
13. Smith PL, Wise RA, Gold AR, Permutt S. Upper airway pressure-flow relationships in obstructive sleep apnea. J Appl Physiol 1988;64:789-95.
14. Kribbs NB, Pack AI, Kline LR, et al. The effects of one night without nasal CPAP treatment on sleep and sleepiness in patients with obstructive sleep apnea. Am Rev Respir Dis 1993;147: 1162-8.
15. Zorick F, Merlotti L, Conway W, Fujita S, Rosenthal L, Roth T. Patient response to various treatments for sleep apnea. Sleep Res 1988;17:270.
16. Anand VF, Ferguson PW, Schoen LS. Obstructive sleep apnea: a comparison of continuous passive airway pressure and surgical treatment. Otolaryngol Head Neck Surg 1991;105:382-9.
17. Sullivan CB, Berthen-Jones M, Issa FG, Eves L. Reversal of obstructive sleep apnea syndrome by continuous positive airway pressure applied through the nares. Lancet 1981;1:862-5.
18. Sanders MH, Moore SE, Eveslage J. CPAP via nasal mask: a treatment for occlusive sleep apnea. Chest 1983;83:144-5.
19. Rappoport DM, Sorkin B, Garay SM, et al. Reversal of the Pickwickian syndrome by the long term use of nocturnal airway pressure. N Engl J Med 1982;301:931-3.
20. Kribbs NB, Pack AI, Kline LR, et al. Objective measurement of patterns of nasal CPAP use by patients with obstructive sleep apnea. Am Rev Respir Dis 1993;147:887-95.
21. Reeves-Hoche HK, Neck R, Zwillich CW. Nasal CPAP: an objective evaluation of patient compliance. Am J Respir Crit Care Med 1994;149:149-54.
22. Katsantonis GP, Schweitzer PK, Branham GH, et al. Management of obstructive sleep apnea: comparison of various treatment modalities. Laryngoscope 1988;98:304-9.
23. Fujita S, Conway W, Zorick F, Roth T. Surgical correction of anatomic abnormalities of obstructive sleep apnea syndrome: uvulopalatopharyngoplasty. Otolaryngol Head Neck Surg 1981; 89:923-7.
24. Sher AE, Schechtman KB, Piccirillo JF. The efficacy of surgical treatments for obstructive sleep apnea. Sleep 1996;19:156-77.
25. Riley R, Guilleminault C, Powell N, Simmons FB. Palatopharyngoplasty failure, cephalometric roentgenograms and obstructive sleep apnea. Otolaryngol Head Neck Surg 1985;93: 240-4.
26. Simmons FB, Guilleminault C, Miles LE. The palatopharyngoplasty operation for snoring and sleep apnea. An interim report. Otolaryngol Head Neck Surg 1984;92:375-80.
27. Piccirillo JF. Outcomes research and otolaryngology. Otolaryngol Head Neck Surg 1994;111:764-9.
28. Sullivan LW. The need for medical treatment effectiveness research. JAMA 1991;226:3264.
29. Elwood PM. Shatuck lecture. Outcome management: a technology of patient experience. N Engl J Med 1988;318:1549-56.
30. Epstein AM. The outcomes movement—will it get us where we want to go? N Engl J Med 1990;323:266-9.
31. Wennberg JE. Outcomes research, cost containment, and the fear of health care rationing. N Engl J Med 1990;323:1202-4.
32. Brook RH, McGlynn EA. Maintaining quality of care. In: Ginzberg E, editor. Health services research. Key to health policy. Cambridge (MA): Harvard University Press; 1991. p. 284-314.
33. Roper WL, Winkenwerder W, Hackbarth GH, Krakauez H. Effectiveness in health care. An initiative to evaluate and improve medical practice. N Engl J Med 1988;319:1197-202.
34. Relman AS. Assessment and accountability. The revolution in medical care. N Engl J Med 1988;319:1220-2.
35. Johns MW. Daytime sleepiness, snoring, and obstructive sleep apnea. The Epworth Sleepiness Scale. Chest 1993;103:30-6.
36. Ware JE, Sherbourne CD. The MOS 36 Item Short-form Health Survey (SF-36): I. Conceptual framework and item selection. Med Care 1992;30:473-83.
37. Measuring functioning and well-being. Durham (NC): University Press; 1992.
38. Ware JE Jr, Snow KK, Kosinski M, Gandek B. SF-36 health survey. Manual and interpretation guide. Boston: The Health Institute; 1993.
39. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-item health survey 1.0. Health Econ 1993;2:217-27.
40. Piccirillo JF, Schechtman KB, Edwards D, Sher AE, White DL. Development of the obstructive sleep apnea quality of life measure [Abstract]. Sleep 1994;Abstract 443.
41. Sher AE, Thorpy MK, Shprintzen RJ, Spielman AJ, Burack B, McGregor PA. Predictive value of the Müller maneuver in selection of patients with uvulopalatopharyngoplasty. Laryngoscope 1985;95:1483-7.
42. Mallampati SR, Gugino LD, Desai SP, et al. A clinical sign to predict difficult tracheal intubation: a prospective study. Can Anaesth Soc J 1985;32:429-34.

43. Kaplan MH, Feinstein AR. The importance of classifying initial co-morbidity in evaluating the outcome of diabetes mellitus. J Chronic Dis 1974;27:387-404.
44. Ewing JA. Detecting alcoholism: the CAGE questionnaire. JAMA 1984;252:1905-7.
45. Kitchens JM. Does this patient have an alcohol problem? JAMA 1994;272:1782-7.
46. American Sleep Disorders Association. 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.
47. Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook PR, Keenan S. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. Sleep 1986;9:519-24.
48. Iezzoni LI. Risk and outcome. In: Iezzoni LI, editor. Adjustment for measuring health care outcomes. Ann Arbor (MI): Health Administration Press; 1994. p. 1-28.
49. Feinstein AR, Wells CK. A clinical-severity staging system for patients with lung cancer. Medicine (Baltimore) 1990;69:1-33.
50. Fleiss J. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley & Sons; 1981. p. 212-36.
51. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297-334.
52. Concato J, Schwartzman D, Feinstein AR. A comparison of logistic regression and conjunctive consolidation as methods of multivariable analysis [Abstract]. Clin Res 1994;42:291a.
53. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care 1990;28:632-42.
54. Kazis LE, Anderson JJ, Meenan RF. Effect sizes interpreting changes in health status. Med Care 1989;27:S178-89.
55. Wennberg JE, Gittelsohn A. Small area variations in health care delivery. Science 1973;182:1102-7.
56. Wennberg JE, Gittelsohn A. Variations in medical small areas. Sci Am 1982;246:120-34.
57. Chassin MR, Kosecoff J, Park RE, et al. Does inappropriate use explain geographic variations in the use of health services? JAMA 1987;258:2533-7.