Sleep, 20(10):835–843

An Instrument to Measure Functional Status Outcomes for Disorders of Excessive Sleepiness

*†Terri E. Weaver, *Andréa M. Laizner, *Lois K. Evans, †Greg Maislin, †§Deepak K. Chugh, *Kerry Lyon, ||Philip L. Smith, ||Alan R. Schwartz, ¶Susan Redline, †‡Allan I. Pack and †§David F. Dinges

**School of Nursing; †Center for Sleep and Respiratory Neurobiology; ‡Pulmonary and Critical Care Division, Department of Medicine; and §Unit for Experimental Psychiatry, Department of Psychiatry, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.; ||Department of Medicine, Division of Pulmonary Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, U.S.A.; and ¶Department of Medicine, Case Western Reserve University and Cleveland Veterans Administration Medical Center, Cleveland, Ohio, U.S.A.*

Summary: This article reports the development of the functional outcomes of sleep questionnaire (FOSQ). This is the first self-report measure designed to assess the impact of disorders of excessive sleepiness (DOES) on multiple activities of everyday living. Three samples were used in the development and psychometric analyses of the FOSQ: Sample 1 (n = 153) consisted of individuals seeking medical attention for a sleep problem and persons of similar age and gender having no sleep disorder; samples 2 (n = 24) and 3 (n = 51) were composed of patients from two medical centers diagnosed with obstructive sleep apnea (OSA). Factor analysis of the FOSQ yielded five factors: activity level, vigilance, intimacy and sexual relationships, general productivity, and social outcome. Internal reliability was excellent for both the subscales ($\alpha = 0.86$ to $\alpha = 0.91$) and the total scale ($\alpha = 0.95$). Test-retest reliability of the FOSQ yielded coefficients ranging from $r = 0.81$ to $r = 0.90$ for the five subscales and $r = 0.90$ for the total measure. The FOSQ successfully discriminated between normal subjects and those seeking medical attention for a sleep problem ($T_{157} = -5.88$, $p = 0.0001$). This psychometric evaluation of the FOSQ demonstrated parameters acceptable for its application in research and in clinical practice to measure functional status outcomes for persons with DOES. Thus, the FOSQ can be used to determine how disorders of excessive sleepiness affect patients' abilities to conduct normal activities and the extent to which these abilities are improved by effective treatment of DOES. **Key Words:** Functional status—Quality of life—Sleepiness—Outcome measures—Sleep disorders.

This article reports the development of a new survey device, the functional outcomes of sleep questionnaire (FOSQ), designed to assess the impact of disorders of excessive sleepiness (DOES) on functional outcomes relevant to daily behaviors and quality of life. No such device has yet been developed for sleep disorders medicine. Although there are numerous sleepiness scales, such as the Stanford sleepiness scale (1) and the Epworth sleepiness scale (2), these measure a patient's sleepiness state or behaviors along various dimensions, but do not measure how the sleepiness affects a person's actual daily ability to function. The FOSQ was developed to accomplish the latter. As third-party payers, health care professionals, and health care administrators attempt to provide cost-effective, quality health care, there is increasing concentration on the balance between economic outcomes and the effects of care decisions on a patient's quality of life (3). Thus, the bottom line for the evaluation of the effective delivery of health care is the patients' perspective on how this care has affected their daily life (3). Extending beyond the focus on mortality or morbidity as outcomes, Ellwood (3) suggested that "The centerpiece and unifying ingredient of outcomes management is the tracking and measurement of function and well-being or quality of life" (p. 1552).

This is no less true for issues surrounding the treatment for DOES where the effectiveness of new treatment modalities centers on their ability to improve the functional status of individuals with these disorders. Functional status, as a measure of quality of life, assesses those activities performed routinely in meeting

Accepted for publication May 1997.
Address correspondence and reprint requests to Terri E. Weaver, Ph.D., University of Pennsylvania School of Nursing, Nursing Education Building, Philadelphia, PA 19104-6096, U.S.A.

basic needs and fulfilling roles (4). Unfortunately, there has been little work on functional status outcomes in sleep disorders medicine in general and in DOES in particular, primarily due to the lack of an instrument designed to specifically capture, from the patient's perspective, the impact of DOES on the performance of daily activities. Although functional status measures exist in health care, there are several limitations posed by the application to DOES of existing generic measures of functional status [e.g. functional status index (5), sickness impact profile (6), and medical outcomes study short form 36 (7)]. First, many of these instruments assume a developmentally based hierarchical relationship among the functional activities surveyed (8,9). This concept may not be valid in sleep disorders where the amount of environmental stimuli required to maintain alertness may be more instrumental in determining whether a task is performed rather than the level of cognitive-neuromuscular development (10). The existence of a hierarchical association among the activities most affected by sleep disorders remains unknown.

Second, generic measures of functional status are generally designed for application in a heterogenous population and therefore selectively assess a wide variety of functional areas (i.e. physical function, social function, psychological function) (11). Although this broad approach is ideal for cross-illness comparisons, it limits the depth of assessment possible in those areas most affected by a specific disorder. This affects the potential variability in functional status scores achievable across a range of illness severity and may inaccurately portray individual levels of functional status (9). For example, similar to the effects of sleep deprivation, DOES appear to affect activities that provide minimal external stimulation or that unmask or even potentiate sleepiness, such as driving or passive vigilance (10,12). Measures that do not systematically address these activities may not provide sufficient data regarding functional status in a homogenous population, such as in those who suffer from DOES due to obstructive sleep apnea (OSA).

For all these reasons, a functional status instrument is needed that specifically targets ways in which DOES impair daily waking activities. This article reports on the development and psychometric properties of the FOSQ, a self-report measure designed to assess the effect of DOES on functional status.

METHODS

Instrument development

The FOSQ is a self-administered paper-and-pencil questionnaire at a fifth-grade reading level (13) that takes approximately 15 minutes to complete. "Functional status" was conceptually defined as those everyday behaviors encompassing the areas of physical, mental, and social functioning in daily life (14). On the basis of this definition, Granger's model of disability (15) was selected to serve as the conceptual framework for the generation of the original 74 items. The use of a conceptual model to guide the formulation of items provides a bountiful and theoretically more meaningful source of items (16,17). It produces a more comprehensive instrument and also reduces the risk of including questions generated from the erroneous interpretation of observed data (17). Granger's model (15) was selected as the basis for the development of the FOSQ because it differentiates pathologically predicted deficits from those behaviors experienced by the individual as a result of impairment or societal influences. The model was particularly applicable to sleepiness-related outcomes where the interplay between physiological and pathological impairment and functional status remains unclear. Further, the disablement model was developed to differentiate deficiencies due to aging from those related to specific organ pathology (18). The original 74 items of the survey reflected Granger's six dimensions (18) of orientation, physical independence, mobility, occupation, social integration, and economic self-sufficiency, as well as several additional daily endeavors that could potentially be affected by DOES, such as sexual activity.

For each activity presented in the first section of the questionnaire, the respondent was presented with a pair of stem questions. The first asked the respondent if they had difficulty performing the identified activity because of being sleepy or tired. The words "sleepy" and "tired" were defined in the instructions as "the feeling that you can't keep your eyes open, your head is droopy, that you want to 'nod off', or that you feel the urge to take a nap. These words do *not* refer to the tired or fatigued feeling you may have after you have exercised". Respondents selected one response from a four-point rating scale: no difficulty, a little difficulty, moderate difficulty, and extreme difficulty. A response alternative was available for respondents to indicate that they did not engage in the activity for reasons other than their sleep disorders. If subjects identified that they had no difficulty performing the activity, they were instructed to skip the next question. If respondents indicated that they had difficulty performing the activity, they then went on to complete the subsequent stem question asking how often they had difficulty performing the activity using a four-point scale: once in a while, some of the time, most of the time, and all the time. This section of the questionnaire was followed by a series of questions requesting respondents to rate how difficult it was to be active during different times of the day and a set of questions inquiring about

the impact of sleepiness on sexual activity. The final group of questions asked respondents to indicate on a six-point scale (0 = never did it, 5 = three or more times a week) how frequently they performed a variety of activities.

Content validity

The conceptual relevance of the FOSQ items was evaluated using a matrix of the six content domains of the theoretical framework discussed above to determine whether the items were distributed among all domains (16,19). Once conceptual relevance was established, the items were categorized into eight subscales: orientation, physical independence, mobility, occupation and economic self-sufficiency, social integration, leisure activities, general activity, and intimate relationships. To establish clinical relevance, the judgment quantification or face validity aspect of content validity was implemented using seven judges (19) with expertise in the areas of functional status instrument development, gerontology, DOES, and sleep deprivation. The judges were asked to rate the clinical relevance of each item and the instrument as a whole to DOES using a four-point ordinal scale (1 = irrelevant, 4 = extremely relevant) (19). The index of content validity was determined by the proportion of items receiving a rating of at least three or four across all judges (19). Items that did not receive this level of endorsement were eliminated from the bank of items. After three rounds, 100% of the judges endorsed each of the 74 items included in the final round with no further suggestions for additions, deletions, or rewording. These final 74 items formed the version of the FOSQ used in the psychometric analysis.

Scoring

A mean-weighted item score was calculated for each subscale using only those activities in which a respondent regularly participated. This method prevented the distortion of the score resulting from missing responses or skipped questions, especially those questions not answered because an individual did not engage in the activity for reasons other than DOES. Subscale scores were totaled to produce a global score. The lower the score, the more dysfunctional the respondent was due to DOES.

The protocol for instrument development was approved by the University of Pennsylvania Committee on Studies Involving Human Beings and the Johns Hopkins University Institutional Review Board. Written informed consent was obtained from all subjects.

Sample and procedure

Three samples were used in the assessment of the validity and reliability of the FOSQ.

Sample 1

The first sample (n = 153) used in the analyses of frequency of endorsement, reliability, and construct and discriminant validity consisted of a convenience sample of 133 individuals recruited on their initial visit to the sleep disorders clinics of the University of Pennsylvania (n = 52) and Johns Hopkins University (n = 81) and a group of normal individuals who did not have a sleep disorder (n = 20). The normal subjects were included to increase the heterogeneity of the sample. Optimally, a heterogenous sample is employed when evaluating the construct validity of a measure to enhance the variability of the data.

Inclusion criteria for the patients in this sample included having at least a seventh-grade education and being studied polysomnographically (PSG). Subject report of excessive daytime sleepiness was not required for participation in the study. Exclusion criteria consisted of a medical history of blindness, a major mental illness (such as depression or dementia), or medical illness (such as heart disease, neuromuscular disease, or obstructive or restrictive lung disease) that would adversely affect the subject's functional status. The demographic characteristics of patients from the University of Pennsylvania and Johns Hopkins University were similar with the exception of educational level ($\chi^2 = 7.67$, $p = 0.02$); more patients from Johns Hopkins University had less than a high school education (Table 1). The mean age of the patients in sample 1 was 49.11 ± 13.04 years. Complete data regarding sample characteristics were available for 84% of the sample. They were predominately male, and 46% were white. Eighty-four percent had at least a high school education, 57% worked full time, and 63% were married. The PSG parameters of the patients in sample 1 included a mean respiratory disturbance index (RDI) of 35.96 ± 32.15, mean rapid eye movement oxygen nadir of 77.22 ± 16.77, and mean body mass index (BMI) of 37.36 ± 7.5.

The normal group included in sample 1 consisted of individuals without DOES who were recruited from the community by advertisement and through personal contact. The exclusion criterion for these individuals was having an increased likelihood of having narcolepsy or OSA determined by history and screening instruments [determined by a value of greater than two on the validated index of narcolepsy-like symptoms and the index for OSA symptoms, as well as a value greater than 0.5 on the multivariable apnea index, pre-

TABLE 1. *Characteristics of subjects in sample 1*

Characteristic[a]	University of Pennsylvania (mean ± standard deviation/%)	Johns Hopkins University (mean ± standard deviation/%)	Normal individuals (mean ± standard deviation/%)
Age	46.52 ± 12.08	51.64 ± 13.69	43.17 ± 8.20
Sex	62% (n = 45) males	66% (n = 73) males	65% (n = 20) males
Race	57% (n = 44) white 41% black 2% other	39% (n = 72) white 49% black 13% other	75% (n = 20) white 25% black
Education	5% (n = 44) < high school 48% high school 48% > high school	22% (n = 72) < high school 31% high school 47% > high school	0% (n = 20) < high school 35% high school 65% > high school
Employed full time	64% (n = 44)	53% (n = 68)	90% (n = 20)
RDI	35.62 ± 29.04	36.24 ± 34.62	
REM O_2 saturation nadir	76.16 ± 13.83	78.03 ± 18.78	
BMI	37.53 ± 7.67	35.25 ± 5.45	24.98 ± 7.02[b]

RDI, respiratory disturbance index; REM, rapid eye movement; BMI, body mass index.
[a] The *t* tests were employed with continuous data and the chi-square statistic with proportional data to test for differences between the patient group (n = 133) and the normal group (n = 20).
[b] $p < 0.0001$.

dictive of OSA, developed by Maislin et al. (20)]. On the basis of this criterion, two of the subjects were excluded from the study, producing a final sample size of 20. These 20 individuals had the following characteristics: mean age of 43.17 ± 8.20, 65% male, 75% white, 100% had at least a high school education, 90% were working full time, and 45% were married. Although the patient and normal groups produced a heterogenous sample with respect to disease, as shown in Table 1, there were no significant differences in demographic characteristics between the patient and normal groups. Given the lack of differences between these two groups with respect to demographic characteristics, their data were combined into one sample for analysis of FOSQ responses. As expected, there were significant differences in BMI between the normal and patient groups ($p < 0.0001$).

After obtaining informed consent, subjects in sample 1 were asked to complete two questionnaires: the subject enrollment form [SEF (20)], which surveys demographic and sleep characteristics, and the FOSQ. For patients, the questionnaires were administered either at the time of the initial visit to the sleep disorders clinic or during the evening of the PSG. Normal subjects completed the questionnaire either at the investigator's office, at the subjects' place of employment, or at home. After completing the questionnaires, these subjects either returned them directly or mailed the completed questionnaires to the investigator.

Sample 2

Sample 2 was comprised of 24 subjects with documented OSA participating in a multisite research project conducted by the University of Pennsylvania, a project that included sites throughout the United States and Canada. This sample was used to determine concurrent validity between the FOSQ and the sickness impact profile (SIP) scale. The mean age of this sample was 41.75 ± 9.14 years; 68% of the subjects were male (n = 20), 52% were white (n = 21), 64% (n = 22) were married, 100% (n = 21) had at least a high school education, and most (86%, n = 21) were employed full time. The mean BMI and RDI were 38.07 ± 7.88 and 50.67 ± 32.71, respectively.

As part of the protocol for another study in which these subjects were participating, written informed consent was obtained followed by completion of the SEF (20), FOSQ, and SIP (16) questionnaires, in that order. The questionnaires were completed in the sleep laboratory prior to the subjects' diagnostic PSG.

Sample 3

Concurrent validity was also evaluated in a sample of 51 subjects with OSA participating in a research project at Case Western Reserve University. Sample 3 was predominantly white (67%), and slightly more than one-half (51%) of the sample was female with a mean age, BMI, and RDI of 49 ± 9.76, 34.94 ± 8.88, and 28.44 ± 24.54, respectively.

Having obtained written informed consent, following their diagnostic PSG, subjects completed the FOSQ and the medical outcomes study short form 36 [SF36 (7)] as part of a battery of measures administered during a day of testing dictated by the protocol of the study in which they were participating.

RESULTS

Frequency of endorsement

Each of the 74 items was evaluated to determine the proportion of subjects that selected each of the re-

sponse alternatives for every item, known as the frequency of endorsement (16). Items containing a response alternative demonstrating a high percentage of selection (>95%) were to be deleted. None of the response alternatives for the items met this criterion so all were retained during this phase of the analysis of content validity.

A review of the frequency of item responses suggested that respondents were responding similarly to the two stem questions: the question about degree of difficulty with an activity and the question regarding the frequency of experiencing difficulty with an activity. In other words, the pattern of responses suggested that these two questions were soliciting the same information. To test this hypothesis, we analyzed the degree of agreement between the responses for these two questions. The proportion of concordant and discordant pairs was determined using the gamma statistic. A gamma of greater than 0.40 was established as the indication of a high level of agreement between pairs. Of the 24 paired questions, 21 (88%) had a gamma value exceeding 0.40. Because the two types of questions appeared to obtain similar information, only one of the pair of stem questions was retained to reduce the length of the questionnaire. This was the stem question asking about difficulty with task performance, which is conceptually the more pertinent inquiry (16). Therefore, 25 questions concerning the frequency of difficulty with daily activities were eliminated from the original 74-item survey.

Initial analysis of internal consistency

Internal consistency, the degree to which each item relates to other items within a scale (21), was determined by correlating each subscale item with the subscale total using Cronbach's coefficient. Inspection of the results of this analysis indicated that 11 questions from the social and leisure subscales depressed the reliability of those subscales. Therefore, these 11 items were deleted, leaving 38 items.

Construct validity

Exploratory factor analysis and internal consistency of the final form of the questionnaire

An exploratory factor analysis using data from sample 1 (n = 153) was used to evaluate the construct validity of the FOSQ. Construct validity assesses the adequacy of an instrument to measure the concept of interest (21). The structure of the FOSQ was determined using the principle components method and varimax orthogonal rotation. Nine factors with eigenvalues greater than 1.00 were identified in the principal factors solution. These factors accounted for a total of 26.77% of the variance. An examination of the scree plot of eigenvalues suggested five factors. This is consistent with the fact that factors 6 through 9 marginally met the standard inclusion criterion of an eigenvalue greater than or equal to 1.00 and individually accounted for less than 2% of the variance. Only one item on these factors loaded highly ($\geq$0.50). Therefore, these four factors were dropped.

A forced five-factor (varimax rotation) solution used 30 of the 38 items from the total scale (Table 2). Items that were dropped did not meet the loading criterion of >0.40. On the basis of the content of the items retained in each factor, factor 1 was labeled "activity level", factor 2 was labeled "vigilance", factor 3 was termed "intimacy and sexual relationships", factor 4 was termed "general productivity", and factor 5 was designated "social outcome". Although the item addressing difficulty with performance of employed or volunteer work had a factor coefficient of 0.39 on the general productivity subscale, it was retained as a component of that factor because it addressed the important issue of work performance and contributed to the internal consistency of that subscale. Because these indices are simple averages of items with factor loadings greater than 0.40, the variance explained is based on an unweighted sum of squares of the factor loadings. These are listed at the bottom of Table 2. Thus, the proportion of total variance in the set of 30 questions captured by the factor structure is approximately equal to the sum of the unweighted final commonality estimates divided by the number of items. This was calculated to be 57.3%. The subscales based on the five factors were considered to be simple summations of the relevant variables (22). Descriptive statistics and Cronbach's alpha internal consistency reliability coefficients for the five-factor–based subscales and the total 30-item FOSQ are presented in Table 3.

Subscale-to-subscale and subscale-to-total correlations

Subscale-to-subscale Pearson correlations ranged from r = 0.52–0.86. This range of intersubscale coefficients indicates the cohesive nature of the measure (23), which is desirable when applying a global score to the evaluation of clinical outcomes. Subscale-to-global FOSQ score intercorrelations ranged from r = 0.78–0.86.

Concurrent validity

Because no functional status instrument for DOES currently exists, two generic measures of functional

TABLE 2. *Factor loadings in the rotated-factor matrix for the FOSQ (n = 159)*

Item	Factor 1	2	3	4	5
Factor 1: activity level					
Difficulty keeping pace with others your own age	0.67	0.14	0.35	0.19	0.20
Difficulty being as active as you want in evening	0.67	0.34	0.23	0.17	0.11
Rating of general level of activity	0.64	0.15	0.31	−0.002	0.08
Difficulty being as active as you want in morning	0.61	0.32	0.25	0.25	0.02
Difficulty exercising or participating in sport activity	0.61	0.08	0.06	0.34	0.16
Difficulty doing work around the house	0.59	0.08	0.07	0.55	0.09
Difficulty doing things for family or friends	0.58	0.17	0.11	0.40	0.25
Difficulty being as active as you want in afternoon	0.52	0.38	0.23	0.24	0.12
Relationship with family/friends been affected	0.45	0.23	0.25	0.22	0.26
Factor 2: vigilance					
Difficulty enjoying theater or lecture	0.24	0.83	0.25	0.16	−0.02
Difficulty enjoying concert	0.15	0.64	0.12	0.20	0.13
Difficulty watching television	0.26	0.55	0.33	0.16	0.28
Difficulty operating motor vehicle for long distances	0.16	0.55	0.14	0.30	0.14
Difficulty watching a movie	0.32	0.52	0.36	0.15	0.30
Difficulty operating motor vehicle for short distances	0.12	0.50	−0.02	0.32	0.11
Difficulty participating in meetings of a group	0.13	0.49	0.36	0.20	0.19
Factor 3: intimacy and sexual relationships					
Ability to become sexually aroused affected	0.32	0.14	0.80	0.14	0.14
Desire for intimacy or sex affected	0.20	0.16	0.79	0.10	0.09
Ability to have an orgasm affected	0.09	0.14	0.72	0.18	0.11
Intimate or sexual relationship affected	0.30	0.24	0.69	0.18	0.09
Factor 4: general productivity					
Difficulty concentrating on things	0.32	0.26	0.24	0.67	0.08
Difficulty taking care of financial affairs and paperwork	0.26	0.30	0.23	0.61	0.13
Difficulty remembering things	0.33	0.06	0.26	0.54	0.23
Difficulty working on a hobby	0.38	0.22	0.08	0.54	0.07
Difficulty finishing a meal	0.01	0.23	0.03	0.45	0.12
Difficulty maintaining telephone conversation	0.11	0.22	0.32	0.45	0.07
Difficulty getting things done because too sleepy to drive	0.37	0.31	0.12	0.40	0.08
Difficulty performing employed or volunteer work	0.35	0.23	0.22	0.39	0.06
Factor 5: social outcome					
Difficulty visiting with family/friends in your home	0.30	0.27	0.30	0.30	0.71
Difficulty visiting with family/friends in their home	0.31	0.35	0.20	0.22	0.66

FOSQ, functional outcomes of sleep questionnaire.
Final unweighted commonality estimates: Factor 1 = 4.65, Factor 2 = 3.75, Factor 3 = 3.67, Factor 4 = 3.52, Factor 5 = 1.59.

status, the SIP and the SF36, were used to examine the concurrent validity of the FOSQ, that is, how the FOSQ related to two external criteria (21). The desirable outcome for this analysis is a moderate relationship between the criterion instruments and the FOSQ, indicating that the FOSQ and the criterion instruments measured the same concept, i.e. functional status, but that the FOSQ provided a unique perspective not captured by the generic measures. A high correlation would indicate that, as an illness-specific measure, the FOSQ did not yield any more information than that obtained by using a generic tool. A low correlation would suggest that the FOSQ did not have concurrent validity.

Sample 2 was used to examine the relationship between the FOSQ and the SIP, and sample 3 was used to examine the relationship between the FOSQ and the SF36. The SIP has well-established psychometric

TABLE 3. *Characteristics of the five-factor–based subscales and total scale of the FOSQ (n = 153)*

Factor	Mean	Standard deviation	Potential range	Obtained range	Cronbach's alpha	Range: item to total correlation
Factor 1: activity level (9 items)	12.41	5.17	0–20	1–20	0.91	0.60–0.76
Factor 2: vigilance (7 items)	12.94	5.05	0–20	2–20	0.87	0.52–0.79
Factor 3: intimacy and sexual relationships (4 items)	14.19	5.78	0–20	0–20	0.89	0.68–0.82
Factor 4: general productivity (8 items)	15.50	4.04	0–20	3–20	0.86	0.82–0.86
Factor 5: social outcome (2 items)	15.57	5.64	0–20	0–20	0.88	0.79
Total scale (30 items)	70.62	21.36	0–120	11.17–100	0.95	0.35–0.73

FOSQ, functional outcomes of sleep questionnaire.

TABLE 4. *Spearman rank order correlation coefficients between FOSQ and SIP, SF36, and the FOSQ global score*

Measure	Global FOSQ	General productivity	Vigilance	Social outcome	Activity level	Intimacy and sexual relationships
SIP (n = 24)						
Overall score	−0.50[a]	−0.74[b]	−0.38	−0.31	−0.57[b]	0.16
Physical dimension	−0.36	−0.48[a]	−0.29	−0.27	−0.33	0.21
Psychological dimension	−0.30	−0.61[b]	−0.23	−0.21	−0.40[a]	0.20
SF36 (n = 51)						
Vitality	0.16[c]	0.19	0.14	0.20	0.32[a]	0.03[c]
Mental health	0.13[c]	0.20	0.05	0.38[b]	0.20	−0.07[c]
General health perception	0.15[c]	0.11	−0.06	0.15	0.19	0.02[c]
Physical functioning	0.28[c]	0.17	0.17	0.14	0.44[b]	0.15[c]
Role emotional functioning	0.46[bc]	0.36[b]	0.14	0.38[b]	0.41[b]	0.26[c]
Role physical functioning	0.26[c]	0.10	0.09	0.20	0.17	0.15[c]
Social function	0.30[c]	0.20	0.19	0.36[a]	0.27	0.12[c]
Bodily pain	0.36[ac]	0.33[a]	0.10	0.33[a]	0.22	0.21[c]

FOSQ, functional outcomes of sleep questionnaire; SIP, sickness impact profile; SF36, Medical Outcomes Study short form 36.
[a] $p \leq 0.05$.
[b] $p \leq 0.01$.
[c] n = 39 for global and intimacy and sexual relationships subscale correlations, because the protocol for completion of the questionnaire did not allow subjects to respond to the questions in the intimacy and sexual relationships subscale if they did not have an intimate or sexual relationship. An intimacy and sexual relationships subscale score is required for computation of the global score.

properties and has been widely used as a measure of functional status outcomes (21). In addition to an overall score, two dimensions, physical and psychosocial, can be obtained. The SF36 is the shortened version of the survey developed by the Medical Outcomes Study (7,21). Eight health concepts are measured by the SF36: physical functioning, role functioning related to physical health problems, role functioning related to emotional problems, mental health, social functioning, vitality (energy and fatigue), bodily pain, and perceptions of general health (7,21).

As displayed in Table 4, significantly reliable correlations were produced between the SIP overall score and the FOSQ global score, indicating that higher functional status as measured by the FOSQ related well with lower disability as determined by the SIP. Comparable measurement of the ability to perform activities of daily living was evident in the significant relationships between the FOSQ general productivity and activity level subscales and the SIP overall score. Likewise, the FOSQ activity level subscale and the SF36 physical functioning subscale were also significantly correlated. Significant moderate correlations were found between the SF36 role emotional functioning subscale and the FOSQ global score and general productivity, social outcome, and activity level subscales. As expected, the FOSQ social outcome subscale was significantly correlated with the SF36 social function subscale but also had a significant association with the SF36 mental health subscale. Activities composing the intimacy and sexual relationships subscale of the FOSQ were not significantly related to any of the SF36 scales.

TABLE 5. *Differences in FOSQ global and subscale scores between patients and normal individuals in sample 1*

	Patient group (n = 133) (mean score ± standard deviation)	Normal group (n = 20) (mean score ± standard deviation)	p value
FOSQ global	68.05 ± 21.24	89.59 ± 8.64	0.0004
Activity level	11.81 ± 4.98	17.51 ± 2.59	0.0001
Vigilance	12.33 ± 5.05	17.43 ± 2.03	0.0007
Intimate and sexual relationships	13.73 ± 5.95	17.54 ± 2.47	0.04
General productivity	15.14 ± 4.04	17.73 ± 3.16	0.04
Social outcome	15.03 ± 5.78	19.39 ± 2.01	0.006

FOSQ, functional outcomes of sleep questionnaire.
To analyze differences between the two groups, *t* tests were used. The Wilcoxon signed rank test was applied when variances did not meet the assumption of homogeneity.

Discriminant validity

A critical evaluation of the utility of a measure is its ability to distinguish between two similar groups that differ on a single characteristic (21). Sample 1 was used to assess differences in FOSQ global and subscale scores between those seeking medical attention for a sleep disorder (n = 133) and normal individuals (n = 20). The analysis yielded significant differences between the patients and normal subjects for FOSQ global and subscale scores (Table 5).

Test-retest reliability

Test-retest reliability was determined by assessing the Pearson correlation of two administrations of the FOSQ using 32 patients from sample 1. These subjects completed the FOSQ both during their initial visit to

the sleep center and again 1 week later, at the time of their scheduled PSG. The analysis was performed on the 30 questions and five subscales yielded by the factor analysis. The correlation coefficient produced by this analysis for the total scale was $r = 0.90$. The test-retest reliability of individual subscales ranged from $r = 0.81$ to $r = 0.90$.

DISCUSSION

The psychometric properties of the FOSQ suggest that it offers a unique self-report measure of functional status as it relates to the impact of DOES on daily activities. The obtained range of scores (Table 3) for the total FOSQ and the five subscales yielded by the factor analysis indicate that the 30-item FOSQ produces a wide range of variability. On the basis of a reliability coefficient of 0.70, suggested by Nunnally and Bernstein (24), the internal consistency reliabilities for the total FOSQ and all subscales shown in Table 3 are acceptable and adequate for the application of the measure in research and practice.

Confirmation that the FOSQ measures functional status impairments was demonstrated by achieving the desired moderate correlations with the SIP and the SF36 for subscales with congruent content. The analysis of the measure's ability to discriminate between normal subjects and those seeking medical attention for a sleep problem suggests that this instrument can identify the functional status impairments imposed by DOES. Finally, the stability of the FOSQ as a measure was demonstrated by acceptable test-retest coefficients.

The five subscales to emerge from the factor analysis of the FOSQ illustrate how the FOSQ yields information about how sleepiness affects a broad range of activities relevant to quality of life. The fact that vigilance surfaced as an area of functional impairment for patients with DOES is not surprising given that the items on this subscale represent the situations associated with the sleepiness prominent in DOES. The items define those functional situations that potentially involve relatively low environmental stimulation and interaction, making it more difficult for the individual to perform the task. This is consistent with the conceptualization of manifest sleepiness (10,25) where the expression of sleepiness is contextually dependent and environmental stimulation plays a key role in masking or unmasking sleepiness (10).

What have previously received little documentation, but were strikingly evident by the results of the factor analysis, were the other functional areas affected by DOES. The emergence of other factors from the factor analysis, such as social outcome and intimate and sexual relationships, provides initial evidence that individuals with DOES encounter problems with aspects of daily living beyond those activities most often identified as affected by sleepiness. Moreover, the emergence of these other impaired functional areas may reflect the ability of the FOSQ to document the impact of conditions other than sleepiness that are associated with some DOES disorders, such as depressed mood or obesity. In this way, the FOSQ, unlike instruments that solely measure sleepiness, captures in a more comprehensive manner the experiences of patients afflicted with DOES.

The relatively high subscale-to-subscale correlations reflect the internal consistency of the measure and indicate the conceptual appropriateness of applying a global score, i.e. the summation of the subscales. The ability of generating an overall singular indication of functional status is desirable in evaluating outcomes in practice and clinical trials.

As outcomes management focuses on methods to improve the patient's quality of life, there is an increased need for instruments that can capture the patient's appraisal of the effectiveness of health care interventions. The FOSQ is the first self-report measure of functional status for DOES with solid psychometric properties that will facilitate the understanding of how DOES intrudes on those activities that make up the quality of daily life. The FOSQ should enable the evaluation of the effectiveness of current and emerging technology to improve the outcome of functional status.

Acknowledgements: The authors thank Jacqueline Fawcett, Ph.D., F.A.A.N., Lorraine Tulman, Ph.D., F.A.A.N., Nancy Kribbs, Ph.D., Joanne Getsey, M.D., and Richard Schwab, M.D., for their contributions to the development of the FOSQ, and Marnie Barrs, R.N., Norman Schubert, R.P.S.G.T., Theresa Roebuck, Andrea Georeno, R.N., and Cynthia Rand, Ph.D., for facilitating data collection. The authors also acknowledge Mark Mahowald, M.D., Gihan Kader, M.D., Thomas Bloxham, M.D., and Charles George, M.D., for their participation in the multisite study that contributed data for this analysis. Support for this research was provided by grants from the Frank Morgan Jones Fund and the Center for Advancing Care in Serious Illness, University of Pennsylvania School of Nursing, Nellcor Puritan Bennett Inc., DeVilbiss, Health Care Inc., Healthdyne Technologies, Respironics, Inc., and the National Institutes of Health, HL42236, 1-R29-HL-53991–01A1, and 5-MO1-RR00040.

REFERENCES

1. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. *Psychophysiology* 1973;10:431–6.
2. Johns MW. Daytime sleepiness, snoring, and obstructive sleep apnea. The Epworth sleepiness scale. *Chest* 1993;103:30–6.
3. Ellwood PM. Shattuck lecture—outcomes management. A technology of patient experience. *N Engl J Med* 1988;318:1549–56.

4. Leidy NK. Functional status and the forward progress of merry-go-rounds: toward a coherent analytical framework. *Nurs Res* 1994;43:196–202.
5. Jette AM. The functional status index: reliability and validity of a self-report functional disability measure. *J Rheumatol* 1987; Suppl. 15:14–9.
6. Bergner M, Bobbitt RA, Carter WB, Gilson B. The sickness impact profile: development and final revision of a health status measure. *Med Care* 1981;19:787–805.
7. Ware JE, Snow KK, Kosinski M, Gandek B. *SF-36 health survey. Manual and interpretation guide.* Boston: The Health Institute, New England Medical Center, 1990.
8. Spector WD, Katz S, Murphy JB, Fulton JP. The hierarchical relationship between activities of daily living and instrumental activities of daily living. *J Chronic Dis* 1987;40:481–9.
9. Kane RL. Commentary: functional assessment questionnaire for geriatric patients—or the clinical Swiss army knife. *J Chronic Dis* 1987;40:95S–8S.
10. Dinges, DF. The nature of sleepiness: causes, contexts, and consequences. In: Stunkard AJ, Baum A, eds. *Perspectives in behavioral medicine: eating, sleeping, and sex.* Hillside, NJ: Lawrence Erlbaum Assoc., 1989:147–79.
11. Schwartz JS, Lurie N. Assessment of medical outcomes. *Int J Technol Assess Health Care* 1990;6:333–9.
12. Roth T, Roehrs T, Carskadon M, Dement W. Daytime sleepiness and alertness. In: Kryger MH, Roth T, Dement WC, eds. *Principles and practice of sleep medicine.* Philadelphia: W. B. Saunders, 1989.
13. Tiger Software, Inc. *RightWriter upgrade with elements of style.* Coral Gables, Florida, 1991.
14. Rubenstein LV, Calkins DR, Greenfield S, et al. Health status assessment for elderly patients. Report of the Society of General Internal Medicine Task Force on Health Assessment. *J Am Geriatr Soc* 1988;37:562–9.
15. Granger CV. A conceptual model for functional assessment. In: Granger CV, Gresham GE, eds. *Functional assessment in rehabilitation medicine.* Baltimore: Williams and Wilkins, 1984: 14–25.
16. Streiner DL, Norman GR. *Health measurement scales. A practical guide to their development and use.* New York: Oxford University Press, 1989.
17. Tulman L, Fawcett J, McEvoy MD. Development of the inventory of functional status-cancer. *Cancer Nurs* 1991;14:254–60.
18. Granger CV. Goals of rehabilitation of the disabled elderly: a conceptual approach. In: Brody SJ, Ruff GE, eds. *Advances in the state of the art.* New York: Springer Publishing Co., 1986: 27–35.
19. Lynn MR. Determination and quantification of content validity. *Nurs Res* 1986;35:382–5.
20. Maislin G, Pack AI, Kribbs NB, et al. A survey screen for prediction of apnea. *Sleep* 1995;18:58–166.
21. Lareau SC, Breslin EI, Meek PM. Functional status instruments: outcome measure in the evaluation of patients with chronic obstructive pulmonary disease. *Heart Lung* 1996;25:212–24.
22. Kim J, Mueller CW. *Factor analysis: statistical methods and practical issues.* Beverly Hills: Sage Publications, 1978.
23. Anastasi A. *Psychological testing.* 4th edition. New York: McMillan, 1976.
24. Nunnally JC, Bernstein IH. *Psychometric theory,* 3rd edition. New York: McGraw-Hill, 1994.
25. Carskadon MA, Dement WC. The multiple sleep latency test: what does it measure? *Sleep* 1982; Suppl. 2:S67–72.