Development of the FOSQ-10: A Short Version of the Functional Outcomes of Sleep Questionnaire

Eileen R. Chasens, DSN, RN[1]; Sarah J. Ratcliffe[2]; Terri E. Weaver, PhD, RN[3]

[1]Assistant Professor, School of Nursing, University of Pittsburgh, Pittsburgh, PA; [2]Assistant Professor of Biostatistics, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, University of Pennsylvania, Philadelphia, PA; [3]Professor of Nursing, Chair, Biobehavioral and Health Sciences Division, School of Nursing, Center for Sleep and Respiratory Neurobiology, Division of Sleep Medicine, Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA

Introduction: The Functional Outcomes of Sleep Questionnaire (FOSQ), has been used in research and clinical practice to measure the impact of daytime sleepiness on activities of daily living. The purpose of this study was to develop a shorter version of the instrument (FOSQ-10) that may be more easily implemented in clinical practice.

Methods: Data from a study of CPAP-treated obstructive sleep apnea (OSA) patients (Sample 1) (n = 155, AHI = 63 ± 31) were used to develop the FOSQ-10. Of the 30 original FOSQ items, 10 questions representing each of the 5 subscales were selected if they had a normal distribution of responses and the largest pre- to post-treatment effect size. The psychometric properties of the instrument were then evaluated with data from a second, independent sample of CPAP-treated OSA patients (n = 51 AH I = 51 ± 28).

Results: Internal consistency of the FOSQ-10 was α = 0.87. Pre-treatment correlations of the 2 scales was r = 0.96. After 3 months of treatment the correlation was r = 0.97 (P < 0.0001). Subscales were also highly correlated at baseline and following treatment. Effect sizes for both instruments were highly correlated and indicated ability to measure meaningful change. Differences were observed between scores on the FOSQ-10 for normal controls and OSA patients.

Conclusions: The FOSQ-10 is a psychometrically strong instrument that performs similarly to the long version. The rapidly completed and easily scored FOSQ-10 shows promise for application in the clinical setting.

Keywords: Functional status, sleepiness, outcome measure, sleep disorders, CPAP, nasal continuous positive airway pressure, sleep apnea, quality of life

Citation: Chasens ER; Ratcliffe SJ; Weaver TE. Development of the FOSQ-10: a short version of the functional outcomes of sleep questionnaire. *SLEEP* 2009;32(7):915-919.

PROBLEMS WITH SLEEP AND SLEEP DISORDERS RESULTING IN EXCESSIVE DAYTIME SLEEPINESS HAVE BEEN SHOWN TO HAVE AN IMPACT ON DAILY ACTIVITIES affecting quality of life. The Functional Outcomes of Sleep Questionnaire (FOSQ)[1] is the gold-standard, disease-specific instrument designed to assess the impact of sleepiness on the ability to conduct daily activities, conceptually defined as functional status, a component of quality of life. The 30-item FOSQ has proven validity and reliability, performing well as an outcome measure in clinical trials.[2-12] However, it may be too long to easily employ in large-scale studies and monitoring efforts[13] as well as clinical practice. A shortened version may facilitate the assessment of functional impairments consequential to excessive daytime sleepiness as well as the evaluation of treatment-associated outcomes. The purpose of this study was to develop a short version of the FOSQ (FOSQ-10) that (1) could explain at least 90% of the variance in the long version; (2) have psychometric properties and generate scores that were comparable to the original FOSQ (FOSQ-30); (3) be able to show comparable change with treatment, demonstrated by effect sizes, with the FOSQ-30.

Submitted for publication June, 2008
Submitted in final revised form March, 2009
Accepted for publication April, 2009
Address correspondence to: Terri E. Weaver, PhD, RN, FAAN, Professor of Nursing, Chair, Biobehavioral and Health Sciences Division, University of Pennsylvania School of Nursing, Claire M. Fagin Hall, 418 Curie Blvd., Philadelphia, PA 19104-6096; Tel:(215) 898-2992; Fax: (215) 573-7492; E-mail: tew@nursing.upenn.edu

METHODS

Instrument Development

Participants in all 3 of the samples used to develop and test the FOSQ-10 were predominantly male, reflecting the representation of this gender in the OSA population. The FOSQ-10 version of the instrument was developed using data (n = 155) (Sample 1) from a multisite study of the effectiveness of 3 months of objectively monitored continuous positive airway pressure (CPAP) treatment in participants with moderate to severe obstructive sleep apnea (OSA) (apnea + hypopnea index (AHI) = 63 ± 31).[14] Inclusion criteria for participation in this study were men and women ages 20 to 60 years; qualitative clinical evaluation of daytime sleepiness by a sleep physician; AHI ≥ 15 events per hour; and a candidate for treatment with CPAP. Persons were excluded if they had a diagnosis of any coexisting sleep disorder by history or polysomnogram; use of sedative or hypnotic medications; shift work; and any physical, cognitive, or psychological condition that might affect the results. As shown in Table 1, the mean age of the participants was 46.3 ± 9.18 years, 89% were male, and mean body mass index was 37.7 ± 8.49 kg/m^2. Participants completed the original FOSQ at baseline and again after 3 months of CPAP treatment. Mean CPAP use was 4.91 ± 2.05 h/night.

The length of the FOSQ-10 was based on several considerations. First, the overall goal was to develop a brief questionnaire easily applied in clinical practice. Second, we wanted the shorter questionnaire to reflect the 5 domains currently in the FOSQ 30: General Productivity, Activity Level, Vigilance, Social Out-

Table 1—Demographic Characteristics of Sample 1, Sample 2, and Normal Controls

Characteristic	Sample 1 (n = 155) Mean ± SD or %	Sample 2 (n = 51) Mean ± SD or %	Normal Controls (n = 57) Mean ± SD or %
Age	46.3 ± 9.18	48 ± 10	43.17 ± 7.54
Sex	89% male	72% male	66% male
BMI	37.7 ± 8.49	33.98 ± 6.18	26.70 ± 6.6
AHI	63 ± 31	51 ± 28	< 5
ESS (baseline)	14.46 ± 4.93	≥ 10	NA
Mean Hours Adherent to CPAP	4.91 ± 2.05 h/night	5.74 ± 1.38 h/night	NA

comes, and Intimate and Sexual Relationships, and maintain the high level of internal consistency found in the FOSQ 30. A length of 10 questions was determined to meet these goals.

The FOSQ-10 was constructed by first examining the distribution of responses among the 4 possible choices for each item on the long version to determine the frequency of endorsement.[15] Those items with evenly distributed pre- and post-treatment responses that also had the highest treatment effect sizes within each subscale, reflecting the greatest magnitude of change, were identified for inclusion in the shorter instrument. The number of items selected from each subscale was dependent upon the length of the subscale (i.e., the longer the subscale, the more items were selected) as well as the strength of the effect size. Consideration was also given to the cultural applicability or commonality of the task assessed. Although the questionnaire contains a response indicating that the patient does not engage in a particular activity, a decrease in the number of items answered would affect the scoring of a shorter form more than a longer version (with several items in each subscale). For example, as shown in Table 2, falling asleep while taking public transportation had a higher effect size than having difficulty remembering. However, not all patients regularly take public transportation and would have to indicate that they do not participate in that activity. Although difficulties with arousal had a larger effect size (0.44) than difficulty with intimacy (0.39), in some cultures asking such explicit questions about sexual functioning is considered inappropriate. Therefore, we chose the question on intimacy, which had the next highest effect size and would be more culturally acceptable, while providing valuable information regarding the impact of sleepiness on sexual activity. We also selected questions that represented the range of activities within a subscale. For example, although difficulty being active in the afternoon had a higher effect size (1.0) than relationships being affected (0.69), we chose to include the latter because we had already included questions about activity level in the morning and in the evening and believed it was more pertinent to include a question regarding relationships, providing a comprehensive assessment, than to include another question regarding activity level. It was also felt that the question regarding activity in the evening would be sufficient to reflect activity level in the latter half of the day. Thus, of the FOSQ-30 questions, 2 questions were selected from General Productivity, 3 from Activity Level, 3 from Vigilance, and one each from the Social Outcome and the Intimacy and Sexual Relationships subscales. As shown in Table 3, the effect sizes of the selected questions ranged from 0.50 to 1.07, indicating a moderate to large treatment effect and clinically meaningful change[16]; except for the question from the Intimacy and Sexual Relationships subscale, which had a smaller effect size (0.39). Further eliminating items from the FOSQ-10 did not produce a Cronbach α greater than that obtained with the 10-item instrument (see below) and may be too short to comprehensively assess daily functioning. A length of 11 or 12 items also did not appreciably improve the internal reliability of the instrument (Cronbach α; data not shown) and would likely impose on the goal of brevity and clinical utility. Effect sizes for the FOSQ-10 and FOSQ-30 for Sample 1 were similar for the 2 instruments as well as the baseline and post-treatment range of scores (see Table 2). Thus, the 10-item scale was sufficient to provide a global assessment of daily functioning yet short enough to be completed rapidly.

Instrument Evaluation

Sample and Procedure

As an outcome measure, according to Kirshner and Guyatt,[17] psychometric evaluation would include analysis of the internal reliability of the instrument and sensitivity to change. We used a second independent sample to evaluate these psychometric properties in the FOSQ-10. The second sample (n = 51, 72% male) consisted of patients (see Table 1) with obstructive sleep apnea (AHI = 51 ± 28) who participated in a randomized study evaluating continuous flexible airway pressure (CFLEX®, Philips Respironics) (n = 26) compared to CPAP (n = 25). The mean age of the participants was 48 ± 10 years with a mean body mass index of 33.98 ± 6.18 kg/m^2. Mean adherence to treatment was 5.74 ± 1.38 h/night; statistically higher in the CFLEX group than CPAP (CFLEX mean hours of use = 6.13 ± 1.35; CPAP mean hours of use = 5.17 ± 1.28, P = 0.02). Participants in the study completed the original FOSQ at baseline and again after 90 days of treatment. Study criteria included adherence to CPAP treatment ≥ 4 h per night; stable medical and psychiatric condition; no evidence of significant upper respiratory symptoms or anatomical abnormality; baseline Epworth Sleepiness Scale (ESS) score ≥ 10; confirmed diagnosis of OSA with AHI ≥ 10/h; and successful titration of CPAP with an AHI of < 5 with positive pressure. Sample 3 (Table 1), who served as normal controls, were individuals (n = 57) recruited from the community who did not have any sleep disorders as determined by polysomnography.

The scoring scheme for the original version of the FOSQ was applied to the shorter instrument. As FOSQ-10 subscales contained only 1 to 3 questions, for measurement precision, it is anticipated that only the Total score would be employed as the metric for the evaluation of the FOSQ-10. To obtain the Total score, a mean-weighted item score was first computed for those

Table 2—FOSQ-30 Baseline and Post-Treatment Range of Scores and Effect Sizes for Treatment Response from Sample 1 for the FOSQ-30 and FOSQ-10 Subscales

FOSQ Subscale	Question Domains	**FOSQ-30 Baseline Range of Scores**	**FOSQ-30 Post Tx Range of Scores**	**FOSQ-30 Effect Size**	**FOSQ-30 Subscale Effect Size**	**FOSQ-10 Subscale Effect Size**
General Productivity	Concentration	2.5–2.9	3.3–3.5	0.83	1.22	0.99
	Getting things done because too sleepy to drive	0.6–1.2	0.90–2.5	0.81		
	Remembering	2.6–3.0	3.09–3.5	0.50		
	Working on hobby	2.6–3.1	3.3–3.7	0.57		
	Taking care of financial affairs & paperwork	2.8–3.3	3.2–3.5	0.33		
	Employed or volunteer work	2.8–3.2	3.1–3.6	0.33		
	Maintain telephone conversation	3.0–3.5	3.4–3.8	0.34		
	Finishing a meal	3.3–3.7	3.5–3.8	0.16		
Activity Level	Activity in evening	2.2–2.6	3.2–3.5	1.1	1.34	1.4
	Activity in morning	2.4–2.8	3.4–3.8	1.0		
	Activity in afternoon	2.4–2.8	3.4–3.8	1.0		
	Relationships affected	2.7–3.1	3.4–3.7	0.69		
	Keeping pace with others	2.7–3.1	3.3–3.7	0.63		
	General level of activity	2.1–2.4	2.6–2.9	0.63		
	Doing things for family	2.9–3.3	3.4–3.8	0.49		
	Doing housework	2.7–3.1	3.2–3.4	0.45		
	Exercising	2.6–3.1	2.8–3.4	0.21		
Vigilance	Watching movies	2.3–3.7	3.2–3.6	0.90	1.25	1.16
	Watch TV	2.3–2.7	3.05–3.5	0.77		
	Enjoy a lecture	2.1–2.5	3.0–3.5	0.76		
	Driving long distance	2.3–2.8	3.1–3.6	0.69		
	Driving short distance	2.7–3.2	3.4–3.8	0.59		
	Participate in meeting	2.4–2.9	3.1–3.5	0.53		
	Enjoy a concert	2.2–2.7	2.6–3.3	0.34		
Social Outcomes	Visit in your home	2.8–3.3	3.4–3.8	0.57	0.91	1.0
	Visit in their home	2.9–3.3	3.4–3.8	0.55		
Intimacy and Sexual Relationships	Arousal	2.9–3.4	3.4–3.8	0.44	0.61	0.72
	Desire intimacy	2.7–3.2	3.2–3.6	0.39		
	Sexual relationships	2.5–3.0	3.0–3.5	0.38		
	Orgasm	3.0–3.5	3.4–3.8	0.30		

subscales with more than one item. This approach prevented the distortion of the score resulting from missing responses. The Total score was derived by calculating the mean of the subscale scores and multiplying that mean by 5. As shown in Table 3, scores on the FOSQ-10 were quite comparable to scores obtained using the long version and a distribution that included the potential range of scores.

RESULTS

Psychometric Properties

The internal consistency of the 10-item instrument was similar to the long form with a Cronbach α of 0.87 (compared to 0.95 for the long version). Prior to treatment, the Total score of

Table 3—Baseline and Post-Treatment Means, Ranges and Correlations for FOSQ-30 and FOSQ-10 Subscales for Sample 2

Scale	Baseline						Post 90-days Follow Up					Treatment Δ	
	Possible Range	Mean	Min	Max	Pearson *r*	*P* Value	Mean	Min	Max	Pearson *r*	*P* Value	Pearson *r*	*P* Value
General Productivity	1-4	2.85 ± 0.70	1.3	4	0.83	< 0.0001	3.58 ± 0.43	2.6	4	0.90	< 0.0001	0.73	< 0.001
FOSQ-10 General Productivity	1-4	2.41 ± 0.92	1	4			3.31 ± 0.71	2.0	4				
Activity Level	1-4	2.31 ± 0.73	1	4	0.91	< 0.0001	3.33 ± 0.54	2.0	4	0.93	< 0.0001	0.91	< 0.001
FOSQ-10 Activity Level	1-4	2.14 ± 0.83	1	4			3.33 ± 0.58	2.0	4				
Vigilance	1-4	2.40 ± 0.73	1	3.9	0.84	< 0.0001	3.38 ± 0.56	1.6	4	0.91	< 0.0001	0.85	< 0.001
FOSQ-10 Vigilance	1-4	2.65 ± 0.73	1	4			3.53 ± 0.51	1.7	4				
Social Outcomes	1-4	2.84 ± 0.86	1	4	0.97	< 0.0001	3.62 ± 0.62	1.0	4	0.95	< 0.0001	0.95	< 0.001
FOSQ-10 Social Outcomes	1-4	2.84 ± 0.86	1	4			3.70 ± 0.61	1.0	4				
Intimacy and Sexual Relationships	1-4	2.54 ± 0.96	1	4	0.89	< 0.0001	3.26 ± 0.85	1.0	4	0.93	< 0.0001	0.89	< 0.001
FOSQ-10 Intimacy and Sexual Relationships	1-4	2.32 ± 1.02	1	4			3.25 ± 0.93	1.0	4				
Total Score	5-20	13.04 ± 3.04	7.03	19.7	0.96	< 0.0001	17.7 ± 2.50	9.5	20	0.97	< 0.0001	0.95	< 0.001
FOSQ-10 Total Score	5-20	12.48 ± 3.23	6	19.3			17.1 ± 2.57	9	20				

the FOSQ-10 was robustly associated with the FOSQ-30 Total score (r = 0.96, P < 0.0001), explaining 92% of the variance in the longer version. The subscale scores of the 2 instruments were also highly related and statistically reliable (see Table 3) with a similar range of pre-treatment scores.

Following CPAP treatment, both the FOSQ-30 and the FOSQ-10 detected a large,[16] clinically meaningful change in the Total score (P < 0.0001). The correlation between the post-treatment FOSQ-10 Total score and the same score for the FOSQ-30 was r = 0.97 (P < 0.0001, 94% of the variance). As illustrated in Table 3, the post-treatment and change scores of the 2 instruments were also highly related.

In evaluating the utility of the FOSQ-10 to distinguish between known groups, we compared scores on the short form for patients with OSA (Sample 2) to normal values for the FOSQ-10.[14] As expected, the baseline Total score on the FOSQ-10 for the OSA group (mean = 12.48 ± 3.23) statistically differed (t = 8.65, P < 0.0001) from the normal values (mean = 17.81 ± 3.10). This suggests that the FOSQ-10 can successfully discriminate between normal individuals who do not experience sleepiness-related impairment from those suffering limitations in daily activities related to excessive daytime sleepiness.

DISCUSSION

The degree to which impairment related to sleepiness is evaluated and the benefit of treatment documented within the context of routine practice depends on the convenience and psychometric strength of outcome measures. We have met the objectives of this study by developing a short, easily applied measure with strong psychometric properties of reliability and validity that adequately assesses how the symptom of daytime sleepiness affects daily activities.

The FOSQ-10 captures the content of the original FOSQ domains and related operational definitions as demonstrated by the high proportion of the variance of the FOSQ Total score explained by the shorter instrument and the robust relationship between the FOSQ-30 and FOSQ-10 subscale pre-treatment and post-treatment scores. The change scores of the two instruments trend together, and although the short form has fewer questions obtaining less information, the strong correlation indicates that similar conclusions can be drawn when using the short form as would be the case if the longer form were employed. Applying the criteria suggested by Nunnally and Bernstein, the internal consistency of the short form surpasses the threshold of 0.70 for application of the measure in both research and clinical practice.[18] The range of pre-treatment and post-treatment scores obtained for the Total score of the short form indicates that, like the original instrument, it has the ability to detect a wide range of functional limitations. The two versions of the FOSQ were able to reach the same statistical conclusion regarding differences between normal controls and patients with OSA.

With similar psychometric performance, the FOSQ-10 provides a simple means to assess functional status both in the clinical arena and in large-scale health assessments with minimizing information loss.[13] The importance of measuring clinically salient outcomes such as functional status has been highlighted by Ellwood who considers functional status, an aspect of quality of life, a key component of outcomes management.[19] Indeed, the American Academy of Sleep Medicine has identified improving quality of life as an indication for treatment.[20]

The fact that the short version of the FOSQ is brief, can be easily administered, and rapidly completed makes it an ideal measure to assess functional status in clinical practice. In this circumstance, only the Total score would be used in interpreting the degree of impairment associated with daytime sleepiness. If a more comprehensive evaluation of daily functioning is desired, such as in research, the longer form is recommended where the subscales can be utilized to characterize the nature of the limitation independently within specific domains as well as provide a global assessment. In conclusion, the FOSQ-10 is a psychometrically strong instrument that shows promise as a valid and reliable measure for the clinical evaluation of disorders of excessive daytime sleepiness and the effectiveness of related interventions.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institutes of Health, National Heart Lung and Blood Institute HL53991 (T. Weaver); P50-HL60287 (A. Pack, G. Maislin, D. Dinges, and T. Weaver), Respironics, Inc., Nellcor Puritan Bennett Inc., DeVilbiss Health Care Inc., and Healthyne Technologies, Inc. We wish to thank Respironics, Inc. for providing the Sample 2 data.

DISCLOSURE STATEMENT

This was not an industry supported study. Dr. Weaver has received research support from Respironics; has consulted for Jazz, Sanofi-Aventis, Apnex Medical, and Cephalon; has had the use of research equipment form Respironics and Protech; and has FOSQ license agreements with Jazz, Sanofi-Aventis. Merck, Sleep Solutions, RTI Health Solutions, NV Organon, and Aspire Medical. The other authors have indicated no financial conflicts of interest.

REFERENCES

1. Weaver TE, Laizner AM, Evans LK, et al. An instrument to measure functional status outcomes for disorders of excessive sleepiness. Sleep 1997;20:835-43.
2. Barnes M, Houston D, Worsnop CJ, et al. A randomized controlled trial of continuous positive airway pressure in mild obstructive sleep apnea. Am J Respir Crit Care Med 2002;165:773-80.
3. Faccenda J, Mackay T, Boon N, et al. Randomized placebo-controlled trial of continuous positive airway pressure on blood pressure in the sleep apnea-hypopnea syndrome. Am J Respir Crit Care Med 2001;163:344-8.
4. Monasterio C, Vidal S, Duran J, et al. Effectiveness of continuous positive airway pressure in mild sleep apnea-hypopnea syndrome. Am J Respir Crit Care Med 2001;164:939-43.
5. Wells RD, Freedland KE, Carney RM, et al. Adherence, reports of benefits, and depression among patients treated with continuous positive airway pressure. Psychosom Med 2007;69:449-54.
6. Blanco J, Zamarron C, Abeleira Pazos MT, et al. Prospective evaluation of an oral appliance in the treatment of obstructive sleep apnea syndrome. Sleep Breath 2005;9:20-5.
7. Dinges DF, Weaver TE. Effects of modafinil on sustained attention performance and quality of life in OSA patients with residual sleepiness while being treated with nCPAP. Sleep Med 2003;4:393-402.
8. Hirshkowitz M, Black J. Effect of adjunctive modafinil on wakefulness and quality of life in patients with excessive sleepiness-associated obstructive sleep apnoea/hypopnoea syndrome: a 12-month, open-label extension study. CNS Drugs 2007;21:407-16.
9. Massie CA, Hart RW. Clinical outcomes related to interface type in patients with obstructive sleep apnea/hypopnea syndrome who are using continuous positive airway pressure. Chest 2003;123:1112-8.
10. Montserrat JM, Ferrer M, Hernandez L, et al. Effectiveness of CPAP treatment in daytime function in sleep apnea syndrome: a randomized controlled study with an optimized placebo. Am J Respir Crit Care Med 2001;164:608-13.
11. Schwartz JR, Hirshkowitz M, Erman MK, et al. Modafinil as adjunct therapy for daytime sleepiness in obstructive sleep apnea: a 12-week, open-label study. Chest 2003;124:2192-9.
12. Steward DL, Weaver EM, Woodson BT. A comparison of radiofrequency treatment schemes for obstructive sleep apnea syndrome. Otolaryngol Head Neck Surg 2004;130:579-85.
13. Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:220-33.
14. Weaver TE, Maislin G, Dinges D, et al. Relationship between hours of CPAP use and achieving normal levels of sleepiness and daily functioning. Sleep 2007;30:711-19.
15. Streiner D, Norman G. Health measurement Scales: a practical guide to their development and use. Oxford: Oxford University Press; 1991.
16. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale: Lawrence Erlbaum Associates; 1988.
17. Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chron Dis 1985;38:27-36.
18. Nunnally J, Bernstein I. Psychometric theory. 3rd vol. New York: McGraw-Hill; 1994.
19. Ellwood. Shattuck lecture-outcomes management. A technology of patient experience. N Engl J Med 1988;318:1549-56.
20. Kushida CA, Littner MR, Hirshkowitz M, et al. Practice parameters for the use of continuous and bilevel positive airway pressure devices to treat adult patients with sleep-related breathing disorders. Sleep 2006;29:375-80.