Performance of Two Self-Report Measures for Evaluating Obesity and Weight Loss*

Donald L. Patrick,† Donald M. Bushnell,‡ and Margaret Rothman§

Abstract

PATRICK, DONALD L., DONALD M. BUSHNELL, AND MARGARET ROTHMAN. Performance of two self-report measures for evaluating obesity and weight loss. *Obes Res.* 2004;12:48–57.

Objectives: To evaluate performance of the Obesity and Weight-Loss Quality-of-Life (OWLQOL) and Weight-Related Symptoms (WRSM) measures.

Research Methods and Procedures: Four studies of obese persons 18 to 75 years of age were analyzed: a 12-week initial validation study, a clinical trial using blinded endpoint data at 50 to 83 weeks, and community studies conducted in the U.S. and Europe. Fifty-six initial validation study subjects visited 1 week after screening to evaluate reproducibility.

Results: Overall, 6107 obese persons completed one assessment, 291 completed follow-up at 12 weeks, and 642 at >50 weeks. Psychometric analyses resulted in a 17-item OWLQOL with a single score tested on five samples that was internally consistent (α values > 0.90) and reproducible (intraclass correlation coefficient > 0.95). The OWLQOL score (higher is better) was associated, as expected, with the symptom measure (lower is better, -0.54), generic quality of life measure (0.53), and measures of physical (0.40) and mental functioning (0.47). The 20-item WRSM was internally consistent ($\alpha = 0.87$) and reproducible (intraclass correlation coefficient = 0.83). The OWLQOL discriminated between genders ($p < 0.001$), presence of disability days ($p < 0.05$), levels of BMI ($p < 0.05$), and levels of symptom bothersomeness ($p < 0.001$). Evaluation at 12 weeks yielded an effect size for ≥2.5% weight loss of 0.77 for the OWLQOL and -0.54 for the WRSM. At ≥50 weeks for ≥10% weight change, effect sizes were 1.63 and -0.73, respectively.

Discussion: The OWLQOL and WRSM are brief, valid, reproducible, and responsive self-reported outcomes for evaluating obesity and weight loss.

Key words: quality of life, reliability, validity

Introduction

Obesity is a major public health problem that increases the risk for comorbid conditions, particularly diabetes, hypertension, coronary artery disease, and cancer (1–5). The prevalence of this condition, defined in adults as a BMI >30 kg/m^2, has been estimated to be between 15% and 20% in industrialized countries (6,7). Overweight and obesity in the United States are increasing and are estimated to affect more than one-half of those over the age of 20 (2,6). Because weight loss has been shown to reduce cardiovascular and other metabolic risk factors (2,8,9), management of obesity is an important health priority. Coping with being overweight and obese and losing or maintaining weight, however, are often significant challenges to individuals not only for personal reasons but also because of the cultural, social, and physical environments that surround them. Dramatic changes in culture, environment, and behavior are warranted.

Increasing BMI has been associated with decreased psychological well-being, reduced social integration, stigmatization, and low self-esteem (10). Obesity also has negative effects on functional status, including work absenteeism, productivity, bodily pain, and depression (11–13).

Patient-reported outcomes, including symptoms, functional status, and perceived quality of life, are increasingly used alongside clinical measures in intervention studies to evaluate weight loss (14). These outcomes include reports of signs and symptoms, impacts on functional status, perceptions of well-being, and evaluations of quality of life. Of the generic functional status instruments, the Short Form

Received for review July 31, 2002.
Accepted in final form October 28, 2003.
*This paper was presented orally at the annual meeting of the International Society for Quality of Life Research, November 7–10, 2001, Amsterdam, The Netherlands, and as a poster at the annual meeting of the North American Association for the Study of Obesity, Oct. 7–10, 2001, Quebec City, Canada.
†Departments of Health Services and Epidemiology, University of Washington, Seattle, Washington; ‡Health Research Associates, Inc., Mountlake Terrace, Washington; and §Johnson & Johnson Pharmaceutical Research and Development, Raritan, New Jersey.
Address correspondence to Donald L. Patrick, Department of Health Services, Box 357660, University of Washington, Seattle, WA 98195-7660.
E-mail: donald@u.washington.edu

36-item health survey (SF-36)[1] has been most widely used (11,15). Generic functional measures, however, do not address key domains relevant to obesity and may not detect minimally important changes to obese persons (16). Obesity-specific measures have been developed, including the 140-item battery constructed from six different psychosocial, health status, and behavioral scales for the Swedish Obesity Study (17), and the Impact of Weight on Quality of Life (IWQOL),[1] a 74-item measure later reduced to the IWQOL-Lite of 31 items (16,18).

Previously developed weight-specific measures were either developed for application to more severely obese populations or without cross-cultural input into the item generation process. Because attitudes toward obesity and weight loss have differing relevance, importance, and sensitivity across different cultures, there is a clear need for measures that not only assess quality of life (QoL) in a broad range of persons who are overweight and obese but also for measures developed with specific and concurrent inclusion of items from multiple cultures.

Also in contrast to previously developed measures, the Obesity and Weight-Loss Quality-of-Life (OWLQOL) was developed using a needs-based theoretical model for perceived QoL that drove identification and selection of items that assess feelings that are unobservable to others, apply to all persons with the condition, are important to meeting the needs of the individual with the condition, and are developed with cross-cultural input (19–24). The OWLQOL items all tap a unitary concept of QoL needs related to being overweight or to losing weight. The OWLQOL and Weight-Related Symptoms Measures (WRSM) also were designed to complete a full battery of patient-reported outcomes employing different concepts and different types of patient-reported outcome measures, including obesity-specific symptoms and QoL, general functional status and well-being, person-specific preference measurement, and disability days (25). The OWLQOL and WRSM are intended to be used together and alongside other patient-reported outcomes of functional status, adherence to diet and treatment, and satisfaction with treatment. The WRSM focuses on symptoms commonly associated with obesity and obesity treatment, and the OWLQOL measures a person's global evaluation of position in life related to weight, weight loss, and weight loss treatment. By using these and other patient-reported outcomes, investigators can address the experience of being overweight and obese and of weight loss on a broad spectrum of issues important to patients, their families, clinicians, regulators, payers, and society in general.

This paper reports the results of analyses from four different studies used to assess the measurement performance of these instruments, including measurement model, internal consistency reliability, reproducibility, construct validity, and responsiveness (longitudinal construct validity). Additional conceptual, methodological, and practical criteria were employed prior and during the development and evaluation of both measures (25–29).

Research Methods and Procedures

Studies and Participants

Data used to evaluate the OWLQOL and WRSM were obtained from four studies: an initial validation sample, blinded data from a trial conducted in the U.S., a U.S. community study, and a European community study. These studies, described in Table 1, included the following:

1. An initial validation study comprised of a convenience sample of obese persons. Participants were recruited through newspaper advertisements and weight loss programs in the Seattle area and from five Wellness Clinics located in Chicago, IL; Raritan, NJ; Kingsport, TN; Spring House, PA; and Cincinnati OH.
2. A clinical trial was conducted to evaluate a product for weight loss among obese persons without a diagnosis of diabetes. Participants ($n = 1282$) with baseline data were used in cross-sectional analyses, and 407 participants with endpoint data were included in analyses of weight change and responsiveness. Blinded endpoint data analyzed included the last assessment available between 50 and 83 weeks.
3. A U.S. community study was drawn from a web-based survey panel designed to represent the U.S. general population. Individuals were selected from this panel based on a BMI of >30 kg/m^2 without comorbidity or a BMI of >27 kg/m^2 with the presence of a comorbidity (type 2 diabetes, hypertension, or high cholesterol).
4. A European community sample included respondents from the United Kingdom, Germany, France, and Italy. All data were collected using a mailed questionnaire with the exception of Italy, where questionnaires were self-administered but delivered and collected by study staff. Sampling was designed to be representative of the country, and a subset of obese persons (BMI $\geq$ 30 kg/m^2) was selected from the community sample. Data from all countries were combined for instrument evaluation.

Participants in all studies provided informed written consent approved by an Institutional Review Board. Weight loss information was provided for the clinical trial participants and those enrolled in formal weight loss programs in the initial validation study. Across all studies, potential participants were excluded if they had been exposed to any

[1] Nonstandard abbreviations: SF-36, Short Form 36-item health survey; IWQOL, Impact of Weight on Quality of Life; QoL, quality of life; OWLQOL, Obesity and Weight-Loss Quality-of-Life; WRSM, Weight-Related Symptoms; PQOL, Perceived Quality-of-Life; PCS, physical component summary scores; MCS, mental component summary scores; ICC, intraclass correlation coefficient; SRM, standardized response mean.

Table 1. Demographic characteristics

Characteristic	Initial validation (n = 340)	U.S. clinical trial (n = 1282)	U.S. community obese population (n = 1478)	European community obese population* (n = 3007)
Age (mean ± SD)	45.4 ± 11.6	44.5 ± 10.7	51.1 ± 13.3	47.8 ± 13.6
Gender [n (%) female]	204 (60.0)	1048 (81.7)	590 (39.9)	1825 (60.7)
Ethnicity [n (96) white]	265 (77.9)	1237 (96.5)	1156 (81.6)	N/A
Marital status [n (%) married]	171 (50.3)	N/A	1015 (69.8)	1421 (70.7)†
Income [n (%) ≥ $50,000 total annual household]	140 (41.2)	N/A	579 (39.2)	‡
Education [n (%) college degree]	265 (77.9)	N/A	455 (30.8)	625 (20.8)
BMI (mean ± SD)	36.3 ± 5.3	37.3 ± 5.2	32.9 ± 4.7	33.6 ± 4.9

* European countries are France, Germany, Italy, and the United Kingdom.
† Marital status was unavailable in the United Kingdom (percentage based on 2010 people).
‡ Median income category: France (41.2% ≥ F150,000); Germany (55.5% ≥ DM42,000); Italy (58.0% ≥ £it30,000,000); United Kingdom (51.3% ≥ £35,000).
N/A, not applicable.

experimental drug or device within 30 days before enrollment; were pregnant or nursing; had gastric restrictive surgery or other surgical procedures designed to cause weight loss; had taken any weight loss medication within 1 month before enrollment; had a history of drug or alcohol abuse within the past 2 years; had a malignancy or a history of a malignancy other than squamous or basal cell carcinomas of the skin; had a history of anorexia nervosa, bulimia, major depression, or panic disorder; were currently receiving psychotropic medication; or had a change in smoking habits within 6 months of the study or who planned to change their smoking habits during the study.

Participants were paid $20 per visit in the initial validation study. Participants in the clinical trial, the U.S. community study, and the European community study were not paid for participation.

Assessments

Patient and demographic characteristics, health status, and QoL data were collected. Core data for all studies included age, gender, height, weight, OWLQOL, WRSM, and the Short Form 36-item health survey. Education, marital status, and income were available for the initial validation and U.S. and European community studies but not for the clinical trial. Longitudinal data were available for the initial validation study and the clinical trial but not for the community studies. Patient characteristics were assessed by clinical staff during in-person visits for the initial validation study and the clinical trial. Data were obtained by self-report for the community sample studies.

The OWLQOL evaluates obesity and trying to lose weight in terms of feelings that are unobservable. Following the psychometric evaluation reported here, the OWLQOL contains 17 items (see Appendix 1). Responses are indicated on a seven-point scale that ranges from 0 ("not at all") to 6 ("a very great deal"). Scores are transformed to a 0 to 100 scale, with higher scores indicating higher obesity-specific QoL. The recall period was 4 weeks.

The WRSM is a 20-item, self-report measure for the presence and bothersomeness of symptoms (Appendix 2). Participants responded either "yes" or "no" as to whether they have experienced the symptom in the previous 4 weeks and then indicated the degree of bothersomeness that having the symptom caused them. The bothersomeness response options are on a seven-point scale and range from 0 ("not at all") to 6 ("a very great deal"). A total score is calculated by summing the bothersomeness scores for each symptom. Total scores range from 0 to 120, with higher scores indicating a higher or worse symptom burden.

Perceived Quality-of-Life (PQOL) is a 20-item, self-report measure that assesses satisfaction with the major categories of fundamental life needs and categories of functioning (30,31). The response scale ranges from 1 (extremely dissatisfied) to 10 (extremely satisfied). This global measure includes domains beyond health status, with explicit evaluation of the environment and satisfaction with life in general. Coefficient α for the overall PQOL was 0.93 in the initial validation study.

The Short Form 36-Item Health Survey is a generic measure of functional status and well-being (32). Eight

domain scores and two summary scores (physical and mental) may be calculated. Only the physical and mental component summary scores (PCS and MCS, respectively) are reported here. Scores are transformed *z* scores, with means set to 50 and SDs set to 10 based on a U.S. normative sample. Each domain is scored from 0 ("poor health") to 100 ("optimal health").

Disability Days is a five-item descriptive measure of self-reported disability and loss of work productivity related to a person's health (33). These items are based on standardized national survey items and have been used in previous obesity studies.

Psychometric Analyses

Psychometric testing of the OWLQOL and WRSM was conducted using standardized procedures and instrument review criteria developed by the Scientific Advisory Committee of the Medical Outcomes Trust (29). Item reduction and development of the measurement model were performed sequentially, first on the initial validation study and then with a randomly selected sample (50%) from the clinical trial population, followed by confirmation with second half of the clinical trial data and the two community studies: one in the U.S. and one in Europe.

Forty-one OWLQOL items were evaluated for inclusion in the final instrument. Initial item reduction was based on respondent understanding of the items and response scale and importance rankings in the item development stage. These findings are reported separately (22). Item reduction criteria for the OWLQOL included 1) items with >5% missing data; 2) items that demonstrated a ceiling effect (>50% of respondents selecting the "not at all" response option, suggesting a high degree of "nonrelevance" or lack of responsiveness); 3) an item-to-total correlation <0.40 (suggesting the item may measure something belonging to a different scale); and 4) an item-to-item correlation >0.70 (indicating redundancy among the individual items). A final instrument containing 17 items was created using these criteria.

Standard descriptive statistics using the SPSS were calculated for each OWLQOL item and the total OWLQOL score to identify ranges and the distributions of response choices (34). Mean, SD, median, and percentage of missing data were also computed for each item. Histograms and box plots were used to determine whether the sample was normally distributed.

Guttman-Cronbach's α was calculated to assess internal consistency, or the degree of association, among the items (35,36). Reproducibility (test/retest reliability) was assessed at 1 week on a subset of the initial validation sample using the intraclass correlation coefficient (ICC) (37).

The factor structure of the 17-item OWLQOL was assessed using principal component analysis with a Promax rotation to allow for expected correlations among the factors. A series of exploratory factor analyses were conducted on multiple data sets. First, an exploratory analysis was conducted on the initial validation study and 50% of the clinical trial sample. The structure was tested on the remaining data sets. Factor correlations of ≥0.70 were considered indicative of the presence of single factor (38).

Convergent validity was evaluated by testing a priori hypotheses about how the OWLQOL should perform in relation to other self-report measures. We expected higher correlations of the OWLQOL with the general QoL measure (PQOL) than the more function-related instrument, the SF-36. We also expected higher correlations of the OWLQOL with the MCS of the SF-36 than with the PCS, based on the overlap between mental health and perceived QoL. For known groups validity, we tested the OWLQOL against independent marker variable, i.e., by comparing scores for participants who were mildly, moderately, or severely overweight (BMI = 27 to 29.9, 30 to 34.9, 35.0 to 39.90, and 40 to 50.0 kg/m^2, respectively), for subjects with low, moderate, and high symptom bother (WRSM tertiles) and presence of disability days ("yes" to have you missed any time from work because of illness in the past 4 weeks?). The OWLQOL score should improve (increase) and the WRSM scores should improve (decrease) as BMI and level of symptom bother decrease, and scores on both instruments should be worse in the presence of disability days. Based on previous literature (16), women were also expected to report lower scores than men. ANOVA was used to evaluate differences between groups, and group differences were identified using Scheffe post hoc procedures.

Responsiveness was reported in terms of the standardized response mean (mean change score/SD of the change score). This effect size statistic was used to identify differences in OWLQOL and WRSM scores associated with weight loss over time (39). For the 12-week validation study, we used ≥2.5% decrease in body weight as a marker for minimally important change based on the short follow-up period and amount of weight loss that could be anticipated with adherence to diet and exercise. For the clinical trial with endpoint data between 50 and 83 weeks, we used ≥10% change in body weight as a minimally important weight loss recommended by the International Obesity Task Force (40).

Respondent burden was addressed in two ways. A small study of 10 obese people (4 men and 6 women), between the ages of 18 and 65 years, self-administered the entire test battery, which was timed for each section. Participants were asked to follow the instructions embedded in the survey and to answer each item to the best of their ability. In addition, respondent reaction to the questionnaire was assessed by observing missing data in the initial validation study.

Results

The majority of subjects were women in all studies except the U.S. community study, were white, and were mar-

ried (Table 1). A large percentage (78%) had some college-level education or were college graduates in the initial validation study compared with 21% to 31% in the community studies. Income levels varied widely across countries, and a high percentage of participants had high household incomes in their respective countries. In the initial study, 291 participants completed the 12-week assessment, and in the clinical trial, 642 participants completed an assessment at 50 to 83 weeks.

Analysis of 41 original items in the OWLQOL questionnaire using the item reduction criteria across the four studies resulted in the removal of 24 items.[2] All OWLQOL results presented in the remainder of this paper are for the 17-item instrument. The sequentially conducted principal component analyses on the initial validation study and random first half of the clinical trial dataset resulted in two factors. The first component contributed 51% of the variance (eigenvalue = 8.7), and the second contributed 8% (eigenvalue = 1.4). The two factors were correlated at 0.70, suggesting that these were highly related. In sequential fashion, these analyses were repeated on the second half of the clinical trial dataset, with the first component contributing 55% of the variance (eigenvalue = 9.4), the second factor contributing 5% (eigenvalue = 1.0), and the two factors correlated at 0.71. When applied to the U.S. and European community studies, the principal components analysis yielded a single factor, contributing 63% (eigenvalue = 10.7) and 60% (eigenvalue = 10.2), respectively. In summary, the principal components analyses on all study samples suggested that a single overall score was appropriate for the OWLQOL and confirmed that the needs-based conceptual model, postulating that all items, if important and applying to all persons and all tapping an unobservable feeling, would form a unitary concept describing obesity-related QoL.[2]

Table 2 shows the mean OWLQOL and WRSM scores by age and gender. In all studies, OWLQOL scores were higher (better) and WRSM scores were lower (better) in men than in women ($p < 0.001$). No clear trends were evident for age groups on the OWLQOL and WRSM in the initial validation study, but OWLQOL scores improved and symptom scores worsened as age increased in the remainder of the studies ($p < 0.05$).

The overall OWLQOL score was internally consistent in all studies (Guttman-Cronbach's $\alpha = 0.93$, initial validation; 0.96, U.S. community; 0.95, European community; 0.94, clinical trial). The ICC, evaluating test–retest reliability on 56 subjects in the initial validation study, was only 0.95. The WRSM was also internally consistent (Guttman-Cronbach's $\alpha = 0.87$) and reproducible (ICC = 0.83) in the initial validation study.

The OWLQOL scores showed stronger associations with the general QoL measure (PQOL) and symptom bothersomeness than with the SF-36 component scores, but these differences were not large (Table 3). Similarly, the association between the OWLQOL and MCS of the SF-36 was higher, but not a great deal higher, than the association with the PCS.

Table 4 shows that the predicted relationships between the OWLQOL total score and measures of BMI, symptom bother, and presence of disability days were generally confirmed. OWLQOL scores decreased as BMI levels increased across all studies. The OWLQOL also discriminated between tertiles of the symptom bother score, decreasing as levels of symptom bother increased ($p < 0.001$). OWLQOL scores also decreased when the person reported having work loss days in the initial and European community studies. These results were not confirmed on the U.S. community study and were not available for the clinical trial.

The mean change in actual weight was 10.7 ± 3.6 kg for the initial validation study and 38.4 ± 8.2 kg in the clinical trial. The correlation between weight change and the OWLQOL score was 0.26 in the initial validation study. Correlation between BMI change and OWLQOL change was -0.09 and between BMI and WRSM change was $+0.09$ in the initial validation study.

Responsiveness of the OWLQOL and WRSM bothersomeness score is shown in Table 5 for different levels of weight increase or decrease. The standardized response mean (SRM) was used as the measure of effect size. Using the cut-point of a $\geq 2.5\%$ decrease in weight over the 12-week initial validation study, the SRM for the OWLQOL was 0.77 and -0.54 for the WRSM. Both the OWLQOL and WRSM scores improved slightly in both studies for patients who increased weight. In the clinical trial, for a $\geq 10\%$ weight decrease, the effect size was 1.38 for the OWLQOL and -0.47 for the WRSM. Effect sizes were smaller for less weight change but remained moderately high for the OWLQOL but not for the WRSM. In general, the OWLQOL proved responsive to weight decrease in the two studies with shorter and longer follow-up.

Mean completion time for the OWLQOL was 5 minutes (range, 3 to 8 minutes), and for the WRSM, mean time to completion was 2 minutes (range, 1 to 4 minutes). Minimal missing data (<0.1%) were observed for all questionnaires in the initial validation study.

Discussion

The goal of this study was to provide psychometric evidence evaluating two newly developed self-report measures specific to obesity and weight loss and to establish

[2] For a technical report on item reduction, factor loadings, and other information on the principal component analyses summarized in this paper, please consult http://www.seaqolgroup.org.

Table 2. OWLQOL and WRSM scores

	Initial validation ($n = 340$)		U.S. clinical trial ($n = 1282$)		U.S. community obese population ($n = 1478$)		European community obese population ($n = 3007$)	
Measure	***N***	**mean (SD)**	***N***	**mean (SD)**	***N***	**mean (SD)**	***N***	**mean (SD)**
OWLQOL								
Total population	340	54.5 (22.0)	1267	51.9 (22.1)	1477	61.5 (21.5)	2952	64.9 (23.7)
Age								
(1) 18 to 44	157	55.7 (21.9)	621	51.2 (21.9)	459	58.9 (23.1)	1259	61.5 (24.4)
(2) 45 to 54	112	53.3 (22.1)	412	49.9 (22.6)	391	59.7 (21.7)	705	65.5 (22.7)
(3) 55+	71	53.7 (22.4)	234	57.4 (21.0)	627	64.6 (19.8)	988	68.9 (22.8)
Scheffe post hoc	No Sig Group Diff			(1) × (3)†		(1) × (3)‡		(1) × (2)*
				(2) × (3)‡		(2) × (3)‡		(1) × (3)‡
Gender								
Male	136	64.7 (20.3)	233	65.0 (20.2)	887	67.7 (17.0)	1161	76.3 (18.8)
Female	204	47.7 (20.5)	1034	48.9 (21.4)	590	52.1 (24.0)	1791	57.6 (23.7)
F		56.7‡		108.5‡		213.1‡		517.2‡
WRSM								
Total population	340	25.5 (18.5)	1161	19.3 (16.7)	1470	17.0 (16.8)	2206	21.0 (18.1)
Age								
(1) 18 to 44	157	23.3 (17.8)	574	17.6 (16.0)	459	15.0 (16.3)	980	18.2 (16.5)
(2) 45 to 54	112	27.2 (19.3)	382	20.9 (17.5)	389	17.4 (17.5)	526	21.6 (18.5)
(3) 55+	71	27.8 (18.4)	205	21.2 (16.7)	622	18.2 (16.7)	700	24.5 (19.3)
Scheffe post hoc	No Sig Group Diff			(1) × (2)†		(1) × (3)‡		(1) × (2)*
				(1) × (3)*		(2) × (3)‡		(1) × (3)‡
Gender								
Male	136	20.3 (15.7)	220	15.2 (14.4)	884	14.5 (15.5)	852	16.9 (16.3)
Female	204	29.0 (19.4)	941	20.3 (17.1)	586	20.7 (18.1)	1354	23.6 (18.7)
F		19.1‡		16.4‡		48.8‡		73.2‡

Note: Higher OWLQOL scores indicate higher levels of condition-specific quality of life; higher WRSM scores indicate greater (worse) symptom severity.
* $p < 0.05$; † $p < 0.01$; ‡ $p < 0.001$.

their appropriateness for further use in evaluating weight loss interventions and observational studies with obese individuals. Recommended minimum values for reliability and validity were exceeded for both the OWLQOL and WRSM instruments (29,41).

Results of the principal components analyses across the four studies and five samples indicated justification for unidimensionality and use of a single score for the OWLQOL. Analyses on the random split halves of the clinical trial dataset and the initial validation study yielded component correlations of >0.70, and the community studies confirmed a single scoring strategy. This result supports the use of the needs-based model of item and instrument development and strongly suggests that a brief specific obesity and weight-loss self-report measure can be used alongside weight loss and other clinical changes in future studies. For functional status and other health-related QoL outcomes, generic measures can be used.

Validity of the observed scores obtained from the OWLQOL and WRSM was confirmed using preidentified logical relationships between the concepts contained in the two new instruments and concepts contained in other, pre-

Table 3. Measurement correlation matrix (initial validation, $n = 340$)

	OWLQOL	WRSM bother	SF-36 PCS	SF-36 MCS
OWLQOL	1.00			
WRSM bother	−0.54	1.00		
SF-36 PCS	0.40	−0.56	1.00	
SF-36 MCS	0.47	−0.40	0.04	1.00
PQOL total	0.53	−0.56	0.43	0.60

Note: all correlations significant at the 0.01 level (two-tailed).

viously used instruments. The pattern of correlations was as predicted. We conclude that the OWLQOL is moderately and positively associated with measures of highly similar constructs, including general QoL, mental and physical well-being, and weight-related symptom bother. The pattern of discrimination between men and women, levels of BMI, and presence of disability days added further evidence of convergent and discriminant validity. These results build on the content validation evaluated in the cross-cultural adaptation of the OWLQOL and WRSM (22).

Both the OWLQOL and the WRSM were responsive to shorter- and longer-term reductions in body weight. That both measures also improved for weight increase may be because of participation in a weight-loss study with diet and exercise recommendations. Further examination of weight increase and QoL will depend on studies of weight gain with and without formal weight loss programs. The observed changes in body weight constitute what might be expected for a 12-week period of formal enrollment in a weight-loss program, that is 2.5%. The OWLQOL and WRSM also performed well in the clinical trial with larger percentages of change in body weight (10%).

Other measures have been developed to assess QoL in overweight and obese patients. Mathias et al. (11) tested the reliability and validity of a modification of existing instru-

Table 4. OWLQOL scores by levels of BMI, symptom bother, and work disability days

	Initial validation		U.S. clinical trial		U.S. community obese population		European community obese population	
	N	mean (SD)	*N*	mean (SD)	*N*	mean (SD)	*N*	mean (SD)
BMI (kg/m^2)								
(1) 27.0 to 29.9	34	59.1 (23.2)	52	62.2 (21.2)	358	71.9 (14.6)	199	76.6 (20.0)
(2) 30.0 to 34.9	125	57.7 (21.7)	445	55.3 (20.8)	681	64.3 (18.3)	1895	67.6 (22.5)
(3) 35.0 to 39.9	103	53.3 (22.8)	398	52.4 (22.5)	284	53.4 (22.5)	597	57.7 (24.2)
(4) 40.0 to 50.0	78	48.7 (20.0)	370	45.8 (21.9)	154	40.0 (25.9)	229	53.8 (25.1)
Scheffe post hoc	(1) × (3)* (2) × (4)*		(1) × (3)* (1) × (4)‡		All groups sig‡		All groups‡ except (3) × (4)	
WRSM (Tertiles)								
Tert. 1 (high bother)	107	69.8 (17.8)	404	63.7 (19.4)	478	71.9 (13.7)	670	77.5 (18.6)
Tert. 2 (moderate bother)	115	54.0 (19.4)	379	54.0 (18.8)	484	64.0 (17.7)	766	65.9 (20.4)
Tert. 3 (low bother)	118	41.0 (18.8)	366	38.3 (20.3)	508	49.1 (24.6)	753	49.9 (22.9)
*F***		66.3‡		165.1‡		177.7‡		316.1‡
Disability days§								
No	194	57.5 (20.9)		N/A¶	517	62.7 (21.2)	1467	66.0 (23.0)
Yes	53	45.6 (20.8)			960	60.8 (21.7)	274	61.3 (23.7)
F		13.6‡				2.3		9.6†

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

§ Have you missed any time from work because of illness in the past 4 weeks?

¶ Data were not available for analysis.

** All post hoc (Scheffe) group comparisons for WRSM were significant ($p < 0.001$).

N/A, not applicable.

Table 5. Responsiveness of the OWLQOL and WRSM score

	OWLQOL change score			WRSM bother change score		
	N	mean (SD)	SRM	*N*	mean (SD)	SRM
Seattle validation (12 weeks)						
Change in weight						
Weight increase	109	4.27 (13.25)	0.32	109	−5.56 (13.90)	−0.40
0 to 2.49% decrease	101	9.32 (11.50)	0.81	101	−6.93 (12.45)	−0.56
≥2.50% decrease	81	11.19 (14.57)	0.77	81	−9.26 (17.24)	−0.54
U.S. clinical trial (>50 weeks)						
Change in weight						
Weight increase	35	8.87 (16.23)	0.55	33	0.12 (20.52)	0.01
0 to 4.99% decrease	75	10.81 (13.96)	0.77	71	−1.11 (14.77)	−0.08
5.00 to 9.99% decrease	109	16.56 (14.04)	1.18	134	−6.63 (12.98)	−0.51
≥10.00% decrease	198	26.16 (16.04)	1.63	230	−11.46 (15.41)	−0.74

SRM, standardized response mean (mean change in score/SD of mean change score).

ments in 417 obese and normal weight individuals. This assessment tool, like the OWLQOL, exhibited acceptable internal consistency and test–retest reliability, good construct validity, and moderate responsiveness to increases or decreases in body weight. It provides a battery of scores, however, using different scoring and weighting procedures for each instrument rather than an overall score.

Kolotkin et al. also developed measures designed to assess health-related QoL in obese patients: the IWQOL questionnaire, and a shorter version of this instrument, IWQOL-Lite (16,42,43). Evaluation of IWQOL indicated test–retest reliabilities averaging 0.75 for single items and 0.89 for scales, with scale internal consistency averaging 0.87. Their study also showed that IWQOL was sensitive to treatment aimed at producing weight loss (16). A second study of IWQOL indicated that total scores correlated highly with other measures of QoL and that subscale scores correlated well with counterparts in the assessment battery. This trial again demonstrated that the instrument was responsive to weight loss (44). In an evaluation of 996 obese patients and controls, reliability of the IWQOL-Lite scales ranged from 0.90 to 0.94 and was 0.96 for the total score. Whereas the IWQOL and IWQOL-Lite are both well-designed, validated, and responsive instruments, their contents may be more relevant for heavier or more obese individuals than for those with moderate obesity, as many of the items on these measures are concerned with mobility and completing simple tasks (e.g., getting up from a chair, climbing stairs) (11). These are also measures of *function* or behavior rather than *perceptions* of body weight and trying to lose weight, which characterize the OWLQOL, and thus, are distinguished from the other measures described here. In addition, the OWLQOL and WRSM were developed with cross-cultural input (22).

Functional status was assessed using the generic SF-36, which permits comparison of effects of weight loss across a large number of different conditions and populations rather than incorporated into the OWLQOL and WRSM (45). The brevity of the 17-item OWLQOL and 20-item WRSM should prove attractive in subsequent applications, particularly clinical trials.

Preliminary knowledge of the effect size for the OWLQOL and WRSM permits estimation of appropriate sample sizes for future studies. The results from the present evaluation also indicate that minimally important differences can be interpreted in terms of score changes related to weight loss. Further responsiveness and interpretation studies will be needed in the context of treatment trials, including studies of weight loss with diet, exercise, and drugs.

Assessing the impact of obesity and weight loss from the perspective of the patient is assuming even greater importance as the incidence of obesity increases in much of the world. Both pharmacological and nonpharmacologic interventions are sought that help overweight and obese persons lose weight. Using measures most relevant to obese people in their everyday lives to evaluate weight loss complements the obviously important goal of shedding pounds and reducing the occurrence of coexisting health conditions. These complementary measures give meaning to weight loss and show how behaviors and feelings are associated with being overweight and with losing weight. These measures may also provide a window into the long-term maintenance of weight

loss, an elusive goal of most weight loss programs. The results presented here indicate that the OWLQOL and WRSM are brief, valid, reproducible, and responsive to weight loss and patients' evaluations of their lives. Both instruments may well be useful for assessing the impact of weight and weight loss programs on obesity-related QoL.

Acknowledgments

This research was funded by research contracts from Johnson & Johnson Pharmaceutical Research and Development, LLC, to Donald L. Patrick and Health Research Associates, Inc. Donald L. Patrick and Donald M. Bushnell have no financial stake in Johnson & Johnson. We thank Mona Martin and Jane Goodman for overseeing the fieldwork and Laura Glauda for recruiting the sites outside Seattle and participating in the conception and conduct of this study. We also thank Dennis Gagnon, Mingliang Zhang, Michael Lean, and two anonymous reviewers for helpful comments on different aspects of this study and manuscript.

References

1. **Zhang R, Reisin E.** Obesity-hypertension: the effects on cardiovascular and renal systems. *Am J Hypertens*. 2000;13: 1308–14.
2. **Lyznicki JM, Young DC, Riggs JA, Davis RM.** Obesity: assessment and management in primary care. *Am Fam Physician*. 2001;63:2185–96.
3. **Melanson KJ, McInnis KJ, Rippe JM, Blackburn G, Wilson PF.** Obesity and cardiovascular disease risk: research update. *Cardiol Rev*. 2001;9:208–9.
4. **Sturm R, Wells KB.** Does obesity contribute as much to morbidity as poverty or smoking? *Public Health*. 2001;115: 229–35.
5. **Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ.** Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med*. 2003;348: 1625–38.
6. **Flegal KM, Carroll MD, Kuczmarski RJ, Johnson CL.** Overweight and obesity in the United States: prevalence and trends, 1960–1994. *Int J Obes Relat Metab Disord*. 1998;22: 38–47.
7. **Seidell JC.** Obesity, insulin resistance and diabetes—a worldwide epidemic. *Br J Nutr*. 2000;83(Suppl 1):S5–8.
8. **Brochu M, Poehlman ET, Ades PA.** Obesity, body fat distribution, and coronary artery disease. *J Cardiopulm Rehabil*. 2000;20:96–108.
9. **Rossner S, Sjöström L, Noack R, Meinders AE, Noseda G.** Weight loss, weight maintenance, and improved cardiovascular risk factors after 2 years treatment with orlistat for obesity. European Orlistat Obesity Study Group. *Obes Res*. 2000;8: 49–61.
10. **Lissner L.** Causes, diagnosis and risks of obesity. *Pharmacoeconomics*. 1994;5(Suppl 1):8–17.
11. **Mathias SD, Williamson CL, Colwell HH, et al.** Assessing health-related quality of life and health state preferences in persons with obesity: a validation study. *Qual Life Res*. 1997; 6:311–22.
12. **Karlsson J, Sjöström L, Sullivan M.** Swedish obese subjects (SOS)—an intervention study of obesity: two-year follow-up of health-related quality of life (HRQL) and eating behavior after gastric surgery for severe obesity. *Int J Obes Relat Metab Disord*. 1998;22:113–26.
13. **Seidell JC.** Societal and personal costs of obesity. *Exp Clin Endocrinol Diabetes*. 1998;106(Suppl 2):7–9.
14. **Patrick DL, Chiang YP.** Measurement of health outcomes in treatment effectiveness evaluations: conceptual and methodological challenges. *Med Care*. 2000;38(9 Suppl):II14–25.
15. **Fontaine KR, Bartlett SJ, Barofsky I.** Health-related quality of life among obese persons seeking and not currently seeking treatment. *Int J Eat Disord*. 2000;27:101–5.
16. **Kolotkin RL, Head S, Hamilton M, Tse CK.** Assessing impact of weight on quality of life. *Obes Res*. 1995;3:49–56.
17. **Sullivan M, Karlsson J, Sjöström L, et al.** Swedish obese subjects (SOS)—an intervention study of obesity. Baseline evaluation of health and psychosocial functioning in the first 1743 subjects examined. *Int J Obes Relat Metab Disord*. 1993;17:503–12.
18. **Kolotkin RL, Crosby RD.** Psychometric evaluation of the impact of weight on quality of life-lite questionnaire (IWQOL-lite) in a community sample. *Qual Life Res*. 2002; 11:157–71.
19. **Hunt SM, McKenna SP.** The QLDS: a scale for the measurement of quality of life in depression. *Health Policy*. 1992; 22:307–19.
20. **Karlsson J, Sjöström L, Sullivan M.** Swedish obese subjects (SOS)—an intervention study of obesity. Measuring psychosocial factors and health by means of short-form questionnaires. Results from a method study. *J Clin Epidemiol*. 1995; 48:817–23.
21. **Hunt SM.** The problem of quality of life. *Qual Life Res*. 1997;6:205–12.
22. **Niero M, Martin M, Glauda L, et al.** A newly designed cross-cultural questionnaire on QOL in obesity. *Qual Life Res*. 1999;8:585.
23. **Karlsson J, Persson L-O, Sjöström L, Sullivan M.** Psychometric properties and factor structure of the Three-Factor Eating Questionnaire (TFEO) in obese men and women. Results from the Swedish Obese Subjects (SOS) study. *Int J Obes Relat Metab Disord*. 2000;24:1715–25.
24. **Niero M, Martin M, Finger T, et al.** Multi-cultural development of two new obesity-specific health-related quality of life and symptom measures: the OWLQOL (Obesity and Weight Loss Quality of Life) instrument and the WRSM (Weight-Related Symptom Measure). *Clin Ther*. 2002;24:690–700.
25. **Patrick DL, Erickson P.** *Health Status and Quality of Life in Health Care Evaluation and Resource Allocation.* New York, NY: Oxford University Press; 1993.
26. **The WHOQoL Group.** The development of the World Health Organization quality-of-life assessment (the WHOQoL). In: Orley J, Kuyken W, eds. *Quality-of-Life Assessment: International Perspectives.* Berlin: Springer Verlag; 1994, pp. 41–57.
27. **National Institutes of Health, National Heart, Lung and Blood Institute.** *Clinical Guidelines on the Identification,*

Evaluation, and Treatment of Overweight and Obesity in Adults. Washington, DC: National Institutes of Health; 1998.
28. **Leidy N, Revicki D, Geneste B.** Recommendations for evaluating the validity of quality of life claims for labeling and promotion. *Value Health*. 1999;2:113–27.
29. **Scientific Advisory Committee of the Medical Outcomes Trust.** Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002;11:193–205.
30. **Patrick DL, Danis ML, Southerland LI, Hong G.** Quality of life following intensive care. *J Gen Intern Med*. 1988;3:218–33.
31. **Patrick DL, Kinne S, Engelberg RA, Pearlman RA.** Functional status and perceived quality of life in adults with and without chronic conditions. *J Clin Epidemiol*. 2000;53:779–85.
32. **Ware JE, Snow KK, Kosinski M, Gandek B.** *SF-36 Health Survey Manual and Interpretation Guide.* Boston, MA: New England Medical Center, The Health Institute; 1993, pp. 7:9–7:11.
33. **Patrick DL, Ramsey SD, Spencer AC, Kinne S, Belza B, Topolski TD.** Economic evaluation of aquatic exercise for persons with osteoarthritis. *Med Care*. 2002;39:413–24.
34. **SPSS.** *Statistical Package for the Social Sciences Base 10.0 User's Guide.* Chicago, IL: SPSS, Inc.; 1999.
35. **Guttman L.** A basis for analyzing test-retest reliability. *Psychometrika*. 1945;10:255–82.
36. **Cronbach LF.** Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297–334.
37. **Deyo RA, Diehr P, Patrick DL.** Reproducibility and responsiveness of health status measures: statistics and strategies for evaluation. *Control Clin Trials.* 1991;12(Suppl 4):142S–58S.
38. **Nunnally JC, Bernstein IH.** *Psychometric Theory*, 3rd ed. New York: McGraw-Hill; 1994.
39. **Cohen J.** *Statistical Power Analysis for the Behavioral Sciences* (revised edition). New York: Academic Press; 1997.
40. **International Obesity Task Force.** *National Institutes of Health Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults.* http://www.iotf.org (Accessed December 12, 2003).
41. **Guyatt GH, Feeny DH, Patrick DL.** Measuring health-related quality of life. *Ann Intern Med*. 1993;118:622–9.
42. **Kolotkin RL, Head S, Brookhart A.** Construct validity of the Impact of Weight on Quality of Life questionnaire. *Obes Res*. 1997;5:434–41.
43. **Kolotkin RL, Crosby RD, Kosloski KD, Williams GR.** Development of a brief measure to assess quality of life in obesity. *Obes Res*. 2001;9:102–11.
44. **Patrick DL, Gagnon DD, Mabe MA, Zagari M.** Assessing the clinical significance of Health-Related Quality of Life (HRQOL) improvement in anemic cancer patients receiving Epoetin Alfa. *Eur J Cancer*. 2003;39:335–45.
45. **Ware JE, Kosinski M.** *The SF-36 Physical and Mental Health Summary Scales: A Manual for User's,* Version 1, 2nd ed. Lincoln, RI: QualityMetric; 2002.

Appendix 1. Obesity and Weight-Loss Quality-of-Life (OWLQOL) 17-Item Questionnaire[3]

1. Because of my weight, I try to wear clothes that hide my shape.
2. I feel frustrated that I have less energy because of my weight.
3. I feel guilty when I eat because of my weight.
4. I am bothered by what other people say about my weight.
5. Because of my weight, I try to avoid having my photograph taken.
6. Because of my weight, I have to pay close attention to personal hygiene.
7. My weight prevents me from doing what I want to do.
8. I worry about the physical stress that my weight puts on my body.
9. I feel frustrated that I am not able to eat what others do because of my weight.
10. I feel depressed because of my weight.
11. I feel ugly because of my weight.
12. I worry about the future because of my weight.
13. I envy people who are thin.
14. I feel that people stare at me because of my weight.
15. I have difficulty accepting my body because of my weight.
16. I am afraid that I will gain back any weight that I lose.
17. I get discouraged when I try to lose weight.

Response scale: 0 = not at all; 1 = hardly; 2 = somewhat; 3 = moderately; 4 = a good deal; 5 = a great deal; 6 = a very great deal.

Appendix 2. The 20-item Weight-Related Symptom Measure (WRSM)[3]

Shortness of breath, tiredness, sleep problems, sensitivity to cold, increased thirst, increased irritability, back pain, frequent urination, pain in the joints, water retention, foot problems, sensitivity to heat, snoring, increased appetite, leakage of urine, lightheadedness, increased sweating, loss of sexual desire, decreased physical stamina, skin irritation.

Response scale: yes/no for frequency; 0 = not at all; 1 = hardly; 2 = somewhat; 3 = moderately; 4 = a good deal; 5 = a great deal; 6 = a very great deal bothersomeness.

[3] The OWLQOL and WRSM can be obtained at http://www.seaqolgroup.org. The instruments are copyrighted and cannot be used without permission.