Neurourology and Urodynamics 34:255–263 (2015)

An Overactive Bladder Symptom and Health-Related Quality of Life Short-Form: Validation of the OAB-q SF

Karin S. Coyne,[1]* Christine L. Thompson,[1] Jin-Shei Lai,[2] and Chris C. Sexton[1]
[1]Evidera, Bethesda, Maryland
[2]Department of Medical Social Sciences and Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, Illinois

Aims: The Overactive Bladder Questionnaire (OAB-q) has demonstrated robust psychometric properties in continent and incontinent OAB patients. However, there is a need for a short-form of this instrument for settings where completing the full OAB-q may be too burdensome. The purpose of this manuscript is to describe the validation of the OAB-q short-form. **Methods:** Three studies were used to derive and validate the OAB-q SF: a 12-week, multicenter, open-label clinical trial of tolterodine ER (N = 865 incontinent OAB [I-OAB]; the "Noble Nested Case-Control" [NCC] study; N = 523 healthy controls; N = 396 OAB); and a test–retest validation study (N = 47). Rasch analysis and confirmatory factor analysis (CFA) were performed to assess the subscale structure, and the psychometric performance of the resulting scales was evaluated. **Results:** Based on the Rasch analysis, 6-items were retained in the OAB-q SF Symptom Bother Scale and 13-items were retained in the HRQL scale. CFAs showed excellent model fit and internal consistency in the study populations. Both scales demonstrated good convergent validity, discriminant validity, internal reliability, reproducibility, and responsiveness to change. The OAB-q SF scales clearly differentiated among I-OAB, C-OAB, and healthy controls. **Conclusion:** The OAB-q SF captures the full spectrum of OAB Symptom Bother and HRQL impact with good reliability, validity, and responsiveness, while being less time-consuming for patients to complete. *Neurourol. Urodynam. 34:255–263, 2015.* © 2014 Wiley Periodicals, Inc.

Key words: OAB; short-form; validation; HRQL

INTRODUCTION

Overactive bladder (OAB) is a highly prevalent, chronic medical condition. It is defined as the symptom complex of urinary urgency, usually accompanied by frequency and nocturia, with or without urgency urinary incontinence (UUI), in the absence of urinary tract infection or other obvious pathology.[1] Recent estimates from the Overactive Bladder on Physical and Occupational Limitations (OAB-POLL) study, a cross-sectional survey conducted among men and women in the US aged 18–70, found that between 8% and 17% of men and 20–30% of women suffer from OAB, depending on the severity cut-point used to define presence of OAB.[2] Both men and women with OAB report considerable bother resulting from their symptoms, embarrassment and inconvenience related to managing their condition, higher rates of depression, lower overall physical and mental health, problems with sleep quality, and decreased levels of sexual activity and enjoyment as compared to those without OAB.[3–6]

Given that OAB is a symptom-based condition, patient-reported outcome (PRO) measures are critical to providing a more thorough understanding of the condition's impact and what patients perceive as important treatment benefits. To be appropriate for evaluation of patients' conditions, PROs must be supported by evidence of their reliability, validity, and responsiveness as well as easy and practical to administer.[7,8] However, given the length of the questionnaires, it may not be feasible to be used in busy clinical settings, such as those in primary care practice, or in clinical trials in which patients must complete multiple instruments. As such, short-forms of PROs have frequently been developed to reduce administrative and respondent burden. While short-forms of PROs do not capture the same level of detail as their initial full-length questionnaires, they tend to have less missing data[9,10] and may be more likely to be incorporated into routine care settings and clinical trials.[11]

The Overactive Bladder Questionnaire (OAB-q),[12] a condition-specific questionnaire which was developed to assess symptom bother and the health-related quality of life (HRQL) impact of OAB, consists of an 8-item Symptom Bother scale and 25 HRQL items that form four subscales (Coping, Concern, Sleep, and Social Interaction). The OAB-q has been shown to be reliable, valid, and responsive to change among both continent and incontinent OAB (I-OAB) patients.[12–16] Although the 33-item OAB-q is not an unusually long questionnaire, clinicians, and researchers may benefit from a short-form that could be used when patient burden is an issue and detailed information on individual HRQL domains is not needed.

Thus, the goal of these analyses was to derive a short-form of the OAB-q from three studies (two clinical samples and the other a community sample) to confirm the psychometric properties of the new OAB-q SF within those datasets. To accomplish this, Rasch analysis, confirmatory factor analysis

Conflict of Interest: Karin S. Coyne, Christine L. Thompson, and Chris C. Sexton are employed by Evidera (formerly a division of United BioSource Corporation), which provides consulting and other research services to pharmaceutical, device, government and non-government organizations. As Evidera employees, they work with a variety of companies and organizations and are expressly prohibited from receiving any payment or honoraria directly from these organizations for services rendered.
Funding Information: Funding for this work was provided by Pfizer. Karin S. Coyne, Chris C. Sexton, and Christine Thompson are employees of Evidera (formerly a division of United BioSource Corporation) who were paid consultants to Pfizer in connection with the development of this manuscript.
Grant sponsor: Pfizer
*Correspondence to: Karin S. Coyne, PhD, MPH, Senior Research Leader, Evidera, 7101 Wisconsin Ave., Suite 600, Bethesda, MD 20814.
E-mail: karin.coyne@evidera.com
Received 11 September 2013; Accepted 16 December 2013
Published online 13 January 2014 in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/nau.22559

(CFA) analyses, and psychometric analyses were performed to assess the OAB-q SF's reliability, validity, and responsiveness.

MATERIALS AND METHODS

Samples

Samples were derived from three studies, which are described below. Institutional Review Board approval was obtained prior to each study's initiation, and all participants provided informed consent before data collection was initiated.

Study 1. Study 1 is a clinical sample where data from a 12-week, multicenter, open-label clinical trial of tolterodine ER in which 865 OAB patients were used for all item analyses. Details and results of this study have been published previously.[17] Measures included the OAB-q, a 3-day micturition diary, which assessed micturition frequency, incontinence episodes, and urgency episodes per 24 hr. Response to treatment was assessed at Week 12 by the Physician Perception of Overall Treatment Benefit, which has a three-level scale (no benefit, a little benefit, and much benefit).

Study 2. Study 2 is a community sample in which analyses to evaluate psychometric properties were performed using data from the "Noble Nested Case-Control" (NCC) study. The NCC dataset contained 523 normal controls and 396 OAB patients.[18] Using a clinically validated computer-assisted telephone interview (CATI), a national telephone survey was conducted to estimate the prevalence of OAB in the United States. Quota sampling methods were used to ensure a representative US population with respect to age, gender, and geographic region. A nested case-control study was conducted among people meeting OAB case criteria and age and gender-matched controls.

In addition to the OAB-q, the CATI survey included questions about socio-demographic characteristics, and other PRO measures, including the Medical Outcomes Study Short-Form 36 (SF-36), which assesses generic health status across eight domains,[19] and measures to assess depression and sleep quality.[18]

Study 3. Study 3 was conducted to assess the reproducibility of the OAB-q and other urinary symptom measures in a clinical sample.[16] A total of 47 participants with urinary urgency/OAB were identified and recruited from five urology clinics. Each patient was scheduled for two assessment visits approximately 2 weeks apart (14 + 4 days). Participants were recruited based on clinical expectations that their bladder conditions would remain stable in the absence of any treatment during the 2 weeks between assessment visits. The reproducibility of the OAB-q SF was assessed in this sample after the SF was derived.

Statistical Analysis

The development schema is shown in Figure 1. Rasch analysis, an item response theory (IRT) modeling method that is commonly used to derive short-forms of patient-reported questionnaires,[9] was conducted in WINSTEPS.[20] CFA was conducted in MPLUS. All other analyses were performed using SAS 8.2. The a priori significance level was fixed at 0.05, and all tests were two-tailed. All statistical tests were considered exploratory and no adjustments were made for multiple test procedures.

OAB-q SF Development: Item Analyses and Confirmatory Factor Analyses

As the dimensionality of the original OAB-q was evaluated previously, we did not repeat the process in this study. Rasch analysis was used to determine which items to retain in the Symptom Bother and HRQL short-form scales using data from Study 1. Criteria used to select items included item location, residual correlations, item-total correlations, fit statistics, ceiling and floor effects, and the scale information function. The unit of measurement (item locations, and self-reported OAB symptom bother and HRQL) reported for all the results in this study is a scaled score (logit), a logarithm measure transformation provided by the Rasch measurement model. Rasch models describe mathematically the relationship between a patient's underlying level of the trait being measured and the location ("difficulty") of the item administered on the same continuum (i.e., OAB). These scaled scores theoretically range from negative infinite to positive infinite. Since the goal of this study was to create a short-form measuring the comparable range of OAB symptoms and impact seen in clinics as the original OAB-q long form, items were selected so that their locations were equally distributed across the OAB continuum.

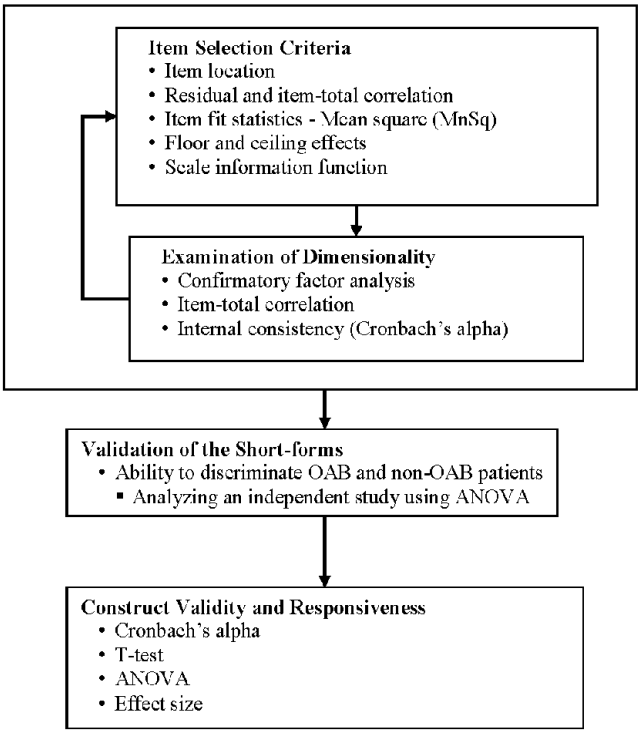

Fig. 1. Short-form development schema.

The relationships among items and between the items and a total score were also examined in selecting short-form items. Item-pairs exhibiting high residual correlations ($\geq$0.5) indicative of local dependency (or redundancy) between items, were candidates for removal from the short-forms. Item-total correlations were also examined in selecting items for the short-forms. For unidimensional instruments, items with relatively low correlations ($<$0.3) were also candidates for removal since they were poorly related to the construct being measured.

Mean-square (MnSq) values were derived from fit statistics from Andrich's[21] rating scale model to determine how well the item fit the measurement model. The ideal MnSq value is 1.0 (i.e., observed variance = predicted variance), but limited unexpected variance is allowed. In this study, because the

items were rated using a polytomous scale, an item with >40% unexpected variance than the model predicted (i.e., infit MnSq values ≥1.4) was considered misfitting.[22]

To determine the extent of floor and ceiling effects in the short-forms, the percentage of patients demonstrating concerns outside the range of the selected items was examined. Lastly, the precision with which the retained items measure OAB along various points in the continuum using the scale information function of the final short-form was assessed. The amount of information provided by an item/scale is related to the precision with which OAB is estimated at that location. Information is affected by: (1) the number of items included in the scale, (2) the quality of scale items, and (3) the match between item location and patient OAB level.[23,24] The shape of the information function curve is more important than its value since its value is determined by the number of items included in the scale. In a general-purpose test, the ideal information function theoretically would be a horizontal line and all subjects would be estimated with the same precision along the continuum defined by the test. Unfortunately, such an information function is almost impossible to achieve in reality. More commonly, the typical shape is a bell-shaped curve in which different OAB levels are estimated with differing degrees of precision. This becomes of considerable importance to both the scale constructor and the consumer since it means that the precision with which a patient's OAB is estimated depends upon where the patient is located on the continuum.[23]

CFA was used to validate the unidimensionality of the two scales following the Rasch analysis using data from Study 1. A hierarchical latent structure was specified for both scales, with two factors for Symptom Bother and three factors for HRQL. Model fit statistics were evaluated, including the comparative fit index (CFI), root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Internal consistency reliability (alphas ≥ 0.70)[25] was evaluated from each study (i.e., study 1 and 2).

Psychometric Performance: Reliability and Validity Analyses. Once the composition and scoring of the instrument was finalized, using data obtained from Study 2, convergent validity was examined by correlating OAB-q SF scores with the CES-D, SF-36 domain scores, and the MOS Sleep scale. Discriminant validity was evaluated by analyzing OAB-q SF scores for the groups C-OAB, I-OAB, and healthy controls via ANOVA with Scheffe's post-hoc comparisons (Study 2).

To evaluate test–retest reliability (reproducibility), intraclass correlations (ICC) and Spearman's correlations were conducted to evaluate the degree of association between scores at Visit 1 and Visit 2 using data from Study 3. Paired *t*-tests were conducted to evaluate whether there were statistically significant score changes between Visit 1 and Visit 2.

Responsiveness analyses were conducted on patients between Visit 1 (screening) and Visit 12 (end of treatment) using data from Study 1. Change scores were calculated, and ANOVA was used to evaluate any differences in OAB-q SF score changes by physician perception of overall treatment benefit.

RESULTS

Sample Characteristics

Socio-demographic characteristics are presented by study and clinical group in Table I. Across studies, most participants were white (range: 84–89%). All participants in Study 1 were diagnosed with OAB with incontinence (I-OAB, n = 865). Participants in Study 2 were categorized in the groups, healthy controls (n = 523), C-OAB (C-OAB; n = 228), and incontinent OAB (I-OAB, n = 168). The mean age ranged from 54 (Study 2, C-OAB) to 66 (Study 3).

Item Analyses

A series of Rasch analyses were performed separately on the Symptom Bother and HRQL subscales, and results for each are discussed in turn. The Symptom Bother short-form best supported by the analyses consisted of six items from the pool of eight items that comprised the original OAB-q Symptom Bother scale. Acceptable fit were found on all items, MnSq <1.4 (i.e., OAB Symptom Bother) from the measurement perspective. Nineteen patients (2.2%) obtained extreme scores, 16 (1.8%) with maximum extremes and three (0.3%) with minimum extremes. These items were reviewed to confirm their clinical relevance. These results indicate the capacity of this short-form to capture the nearly full range of OAB symptom bother defined by this sample.

The separation index of the short-form was 1.73, suggesting these items could discriminate samples into two statistically significant strata. The scale information function curve along the OAB symptom bother continuum is shown in Figure 2. As expected, the highest information function is close to the midpoint of the continuum (scaled scores = 0). Patients' degrees of symptom bother can be precisely measured (i.e., SE < 0.5, where corresponding r > 0.7) when their scaled scores range between −1.92 and 1.92.[26,27] In the current sample, 89.1% (771/865) were within this precision range.

TABLE I. Sample Socio-demographic Characteristics by Psychometric Validation Study

	Study 1	Study 2			Study 3
Variable	I-OAB (N = 865)	Controls (N = 523)	C-OAB (N = 228)	I-OAB (N = 168)	OAB patients (N = 47)
Gender (n, % female)	636 (73.5%)	345 (66.0%)	82 (36.0%)	128 (76.2%)	35 (74.5%)
Age, mean (SD)	61.0 (14.7)	52.1 (16.3)	54.3 (16.7)	60.0 (15.1)	66.0 (12.9)
Race, n (%)					
White	772 (89.2%)	442 (84.5%)	191 (83.8%)	143(85.1%)	40 (85.1%)
African American	72 (8.3%)	37 (7.1%)	20 (8.8%)	15 (8.9%)	5 (10.6%)
Asian	11 (1.3%)	3 (0.6%)	3 (1.3%)	0 (0.0%)	0 (0.0%)
Hispanic	0 (0.0%)	13 (2.5%)	5 (2.2%)	6 (3.6%)	0 (0.0%)
Other[a]	7 (0.8%)	23 (4.4%)	9 (4.0%)	4 (2.4%)	2 (4.3%)
Not reported	3 (0.3%)	5 (1.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

[a]Includes American Indian and Cuban.

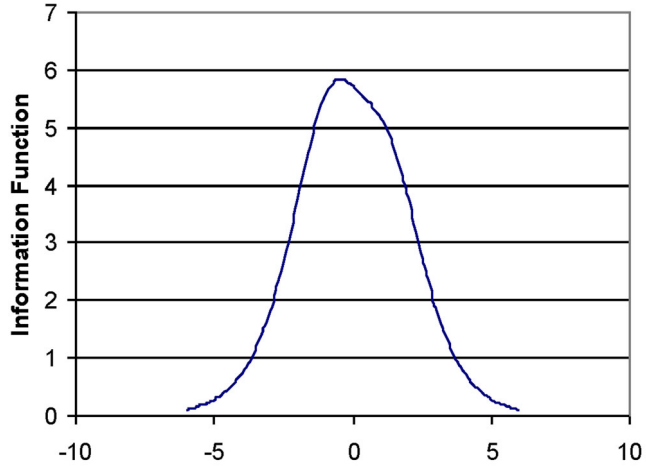

Fig. 2. Scale information function curve: 6-item OAB-q SF Symptom Bother.

Results of the series of Rasch analyses for the HRQL scale pointed to the retention of 13 items from the pool of 25 items that comprise the original OAB-q. All items showed acceptable fit, MnSq <1.4 (OAB HRQL) from the measurement perspective. Twelve patients (1.4%) obtained extreme scores, 7 maximum extremes (0.8%) and 5 with minimum extreme (0.6%). These minimal and negligible ceiling and floor effects indicate the capacity of this short-form to capture the nearly full range of OAB HRQL defined by this sample. The 13-item HRQL short-form has a separation index of 2.57 suggesting these items could discriminate these samples into three statistically significant strata based on their OAB HRQL (e.g., mild, moderate, and severe). The scale information function curve along the OAB HRQL continuum is shown in Figure 3.

As expected, the highest information function appears in the mid-point of the continuum (scaled scores = 0). Patients' OAB related HRQL could be precisely measure (i.e., SE < 0.5, where corresponding r > 0.7) when their scaled scores range between −2.1 and 2.1; 94.7% (819/865) were within this precision range.

Confirmatory Factor Analyses

CFAs were performed to assess the model fit for each subscale using data from Study 1. Both the Symptom Bother scale and HRQL Subscales fit the data well, with Bentler's Comparative Fit Indices (CFI) of 0.953 and 0.976, respectively (Table II). RMSEA values were 0.131 and 0.062. Both the Symptom Bother and HRQL subscale were fitted as hierarchical latent models to achieve the best model fit (two latent factors for Symptom Bother and three for HRQL) but perform well with strong internal consistency when scored as unique scales using classical test theory. Cronbach's alphas for six items comprising the Symptom Bother scale were 0.82 and 0.91 for Study 1 and Study 2 (Table II). The 13-item HRQL subscale also demonstrated excellent internal consistency (Cronbach's alpha: Study 1, 0.92, Study 2, 0.95; Table II).

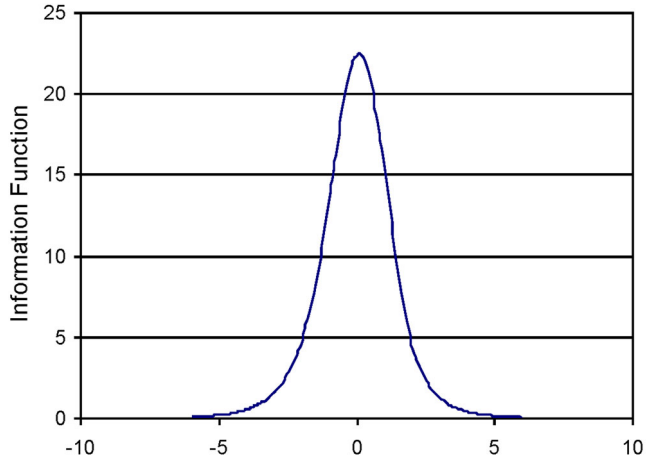

Fig. 3. Scale information function curve: 13-item OAB-q SF HRQL.

The final OAB-q SF instrument and scoring manual are available in Appendix A. Raw scores are then transformed to a 0–100 scale using the formulas provided in the scoring manual. As with the OAB-q, this will provide Symptom Bother scores where higher score values are indicative of greater symptom bother and lower scores indicate minimal symptom bother. For the HRQL scores, higher scores are indicative of better HRQL and lower scores indicate worse HRQL. It is recommended that if <50% of the scale items are missing, the scale should be retained with the mean scale score of the items present used to impute a score for the missing items. If ≥ 50% of the items are missing, no scale score should be calculated, the subscale score should be considered missing.

Psychometric Performance: Reliability and Validity Results

Statistically significant correlations were found between the OAB-q SF scales and other PRO instruments, including the CES-D, SF-36 domains, and the MOS Sleep Scale. As expected, correlations were generally small to moderate, and all were in the direction anticipated. For example, higher levels of depressive symptoms as measured by the CES-D were linearly related to higher levels of Symptom Bother ($r = 0.31$, $P < 0.0001$) and lower levels of HRQL on the OAB-q SF ($r = -0.35$, $P < 0.0001$). Similar correlations were found for these scales in relation to the SF-36 domains (r range −0.27 to 0.37 for Symptom Bother and 0.32 to 0.43, all P values <0.0001). Consistent with hypotheses, correlations tended to be more modest in relation to domains of the MOS Sleep scale, with stronger correlations found for the Sleep Index (HRQL, $r = 0.40$; Symptom Bother, $r = -0.45$; $P < 0.0001$).

Both the Symptom Bother and HRQL scales easily differentiated between Control, C-OAB and I-OAB patients ($P < 0.0001$; Table III). As expected, mean (SD) scores indicated the most impairment for those with I-OAB (Symptom Bother, 48.9 [24.5]; HRQL, 66.8 [23.7]), followed by C-OAB (Symptom Bother, 24.5 [17.6]; HRQL, 85.0 [16.3]), with scores indicating almost no impairment for controls (Symptom Bother, 9.8 [11.7], HRQL 95.4 [8.8]). These results provide evidence that the Symptom Bother and HRQL short-forms developed in this study are valid and can be used in a clinical or research context to differentiate OAB patients from normal, healthy individuals.

TABLE II. Confirmatory Factor Analysis: Model Fit Statistics for OAB-q SF

Statistic	Symptom Bother	HRQL
CFI	0.953	0.976
RMSEA	0.131	0.062
SRMR	0.162	0.033

CFI, comparative fit index; RMSEA, root mean square error of approximation; SRMR, standardized root mean residual.

TABLE III. Discriminant Validity: OAB-q SF Scale Scores for OAB Patients and Controls (Study 2)

OAB-q SF Subscales	Controls	C-OAB	I-OAB
Symptom Bother (mean, SD)[a]	n = 508	n = 222	n = 163
	9.8 (11.7)	24.5 (17.6)	48.9 (24.5)
HRQL (mean, SD)[a]	n = 516	n = 226	n = 163
	95.4 (8.8)	85.0 (16.3)	66.8 (23.7)

[a]Means for the three groups were significantly different from each other at $P < 0.0001$ as indicated by ANOVA with Scheffe's post-hoc comparisons.

The reproducibility of the OAB-q SF was strong as paired *t*-tests of data from Study 3 showed that there were no statistically significant changes in Symptom Bother or HRQL subscale scores from Visit 1 to Visit 2 (Table IV). ICCs were 0.81 and 0.92 for Symptom Bother and HRQL, respectively, and Spearman's correlations were 0.83 and 0.93 for these scales (all significant at $P < 0.001$).

Regarding responsiveness, both OAB-q SF subscales were sensitive to treatment benefit as evaluated by their physicians. All pairwise comparisons using Scheffe's test of multiple comparisons were significant ($P < 0.0001$; Table V). Further examination of the mean change over time in the total sample from baseline to Week 12 showed that participants reported a mean (SD) reduction of 22.5 (21.8) points on the Symptom Bother scale and a mean (SD) improvement of 23.3 (21.7) on the HRQL scale, which translated into effect sizes of −1.14 (Symptom Bother) and 0.99 (HRQL).

DISCUSSION

This study reports on the development and psychometric properties of the OAB-q SF—a PRO instrument to evaluate OAB, which consists of 19 items comprised by two scales: a 6-item Symptom Bother scale and a 13-item HRQL scale. Validation findings presented here were conducted in clinical and community samples, which included patients with continent and I-OAB as well as healthy controls. The demonstration of good psychometric properties in this diverse representation of participants highlights the ability of the OAB-q SF to capture a wide spectrum of OAB symptoms.

This study used the Rasch measurement model, a family member of IRT models, to select items that best covered the full range of the continuum of OAB symptom bother and HRQL impact. This application of IRT is well supported in the literature for promoting instrument precision and overcoming the problem of instrument concentration around the middle of the hierarchy with relatively fewer items positioned at the ends.[28–30] Findings from the Rasch analysis in this study, coupled with CFA results replicated in a second clinical sample, support the development of the 6-item Symptom Bother scale and 13-item HRQL scale. Both scales have considerable relevance to patients and clinical practice. A growing body of research has documented that symptoms of OAB can be highly burdensome to patients[3,6,31] as well as incur considerable economic impact—with annual costs in the US estimated to be over $12 billion.[32]

Additional psychometric analyses presented here documented the internal consistency, reproducibility, convergent validity, discriminant validity, and responsiveness of the OAB-q SF. The convergence of these findings—conducted in three distinct clinical samples—is a strength of this study. Both scales discriminated between patients with I-OAB, C-OAB, and healthy controls, suggesting that the OAB-q SF could be used in a clinical or research context to differentiate OAB patients from normal, healthy individuals. Importantly, while the OAB-q and OAB-q SF were developed for use among OAB patients—including those with UUI and MUI—these instruments were not designed to discriminate between SUI and MUI patients. Data from an observational study did demonstrate that the OAB-q long form detected differences in both Symptom Bother

TABLE IV. Test-Retest Reliability: Comparison of OAB-q SF Scores at Visit 1 and Visit 2 (Study 3)

Subscale	Visit 1 mean (SD)	Visit 2 mean (SD)	Difference score	T-value	Intraclass correlations	Spearman's correlations
Symptom bother	56.2 (24.8)	50.8 (22.5)	−5.4	−2.3	0.81	0.83***
HRQL	56.6 (24.6)	58.6 (24.0)	2.0	1.3	0.92	0.93***

Paired *t*-tests comparing responses at Visit 1 and Visit 2 were nonsignificant.
P-values are: *<0.05, **<0.01, ***<0.001.

TABLE V. Responsiveness: OAB-q SF Change Scores From Visit 1 to Visit 12 (Study 1)

	Physician perception of overall treatment benefit				
OAB-q SF subscales	No improvement mean (SD)	Improved a little mean (SD)	Improved very much mean (SD)	Overall *F*-value	*P*-value
Symptom Bother, mean (SD)	(n = 64)	(n = 247)	(n = 287)	50.67	<0.0001
	−3.7 (17.6)	−18.9 (20.0)	−29.7 (20.7)		
HRQL, mean (SD)	(n = 75)	(n = 308)	(n = 472)	63.63	<0.0001
	4.9 (16.4)	18.2 (19.5)	29.6 (21.2)		

All pairwise comparisons between means were performed using Scheffe's test of multiple comparisons were significant: 1, improved very much versus Improved a little; 2, improved very much versus no improvement; 3, improved a little versus no improvement; *<0.05, **<0.01, ***<0.001.

and the HRQL subscales among SUI, UUI, and MUI patient groups.[33] Future research is needed to determine if the OAB-q SF performs similarly in detecting differences among these patient groups.

The use of anchor-based methods showing that both OAB-q SF subscales were sensitive to treatment benefit as evaluated by their physicians coupled with the robust effect sizes found for change over the treatment period (−1.14 for Symptom Bother; 0.99 for HRQL) provide a strong foundation of support for the responsiveness of the instrument according to current recommendations.[8] However, a limitation of this study is that these data are from one single-arm observational study. Additional data from treatment studies are needed to provide further information about the instrument's sensitivity to change and to estimate the minimally important difference (MID).

The use of standardized PRO questionnaires in clinical practice has been shown to improve communication between patients and physicians and may improve the process of treatment selection.[34–36] In order for a PRO to be widely used and effective, it must be brief, easy to complete, and appropriate for the setting in which it is being used, as well as possess reasonable estimation precision along the construct being measured.[37,38] The OAB-q SF fulfills these criteria and is a promising tool for settings in which completing the full 33-item OAB-q may be too burdensome for patients and physicians.

CONCLUSION

Findings presented here provide the first evidence supporting the psychometric properties of the OAB-q SF. The OAB-q SF is an economical, efficient alternative to longer questionnaires that can be used to evaluate Symptom Bother and HRQL impact in routine clinical care as well in research settings with minimal participant burden. Future research is needed to evaluate the tool's responsiveness.

REFERENCES

1. Haylen BT, de Ridder D, Freeman RM, et al. An International Urogynecological Association (IUGA)/International Continence Society (ICS) joint report on the terminology for female pelvic floor dysfunction. Neurourol Urodyn 2010;29:4–20. Epub 2009/11/27.
2. Coyne KS, Sexton CC, Bell JA, et al. The prevalence of lower urinary tract symptoms (LUTS) and overactive bladder (OAB) by racial/ethnic group and age: Results from OAB-POLL. Neurourol Urodyn 2012;32:230–7. Epub 2012/08/01.
3. Benner JS, Becker R, Fanning K, et al. Bother related to bladder control and health care seeking behavior in adults in the United States. J Urol 2009;181:2591–8. Epub 2009/04/21.
4. Coyne KS, Sexton CC, Kopp ZS, et al. The impact of overactive bladder on mental health, work productivity and health-related quality of life in the UK and Sweden: Results from EpiLUTS. BJU Int 2011;108:1459–71. Epub 2011/03/05.
5. Coyne KS, Sexton CC, Thompson C, et al. The impact of OAB on sexual health in men and women: Results from EpiLUTS. J Sex Med 2011;8:1603–15. Epub 2011/04/16.
6. Kannan H, Radican L, Turpin RS, et al. Burden of illness associated with lower urinary tract symptoms including overactive bladder/urinary incontinence. Urology 2009;74:34–8. Epub 2009/05/12.
7. FDA. Guidance for Industry: Patient-reported outcome measures—Use in medical product development to support labeling claims. In: Office of the Federal Register, National Archives and Records Administration. *Federal Register*. Washington, DC: Food and Drug Administration; 2009. 65132–3.
8. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102–9. Epub 2008/01/08.
9. Cole JC, Rabin AS, Smith TL, et al. Development and validation of a Rasch-derived CES-D short form. Psychol Assess 2004;16:360–72. Epub 2004/12/09.
10. Stroud MW, McKnight PE, Jensen MP. Assessment of self-reported physical activity in patients with chronic pain: Development of an abbreviated Roland-Morris disability scale. J Pain 2004;5:257–63. Epub 2004/06/29.
11. Pincus T, Wolfe F. Patient questionnaires for clinical research and improved standard patient care: Is it better to have 80% of the information in 100% of patients or 100% of the information in 5% of patients? J Rheumatol 2005;32:575–7. Epub 2005/04/01.
12. Coyne K, Revicki D, Hunt T, et al. Psychometric validation of an overactive bladder symptom and health-related quality of life questionnaire: The OAB-q. Qual Life Res 2002;11:563–74. Epub 2002/09/11.
13. Coyne KS, Matza LS, Thompson C, et al. The responsiveness of the OAB-q among OAB patient subgroups. Neurourol Urodyn 2007;26:196–203. Epub 2006/10/04.
14. Coyne KS, Matza LS, Thompson CL. The responsiveness of the Overactive Bladder Questionnaire (OAB-q). Qual Life Res 2005;14:849–55. Epub 2005/07/19.
15. Coyne KS, Matza LS, Thompson CL, et al. Determining the importance of change in the overactive bladder questionnaire. J Urol 2006;176:627–32 ; discussion 32. Epub 2006/07/04.
16. Matza LS, Thompson CL, Krasnow J, et al. Test–retest reliability of four questionnaires for patients with overactive bladder: The overactive bladder questionnaire (OAB-q), patient perception of bladder condition (PPBC), urgency questionnaire (UQ), and the primary OAB symptom questionnaire (POSQ). Neurourol Urodyn 2005;24:215–25. Epub 2005/03/05.
17. Siami P, Seidman LS, Lama D. A multicenter, prospective, open-label study of tolterodine extended-release 4 mg for overactive bladder: The speed of onset of therapeutic assessment trial (STAT). Clin Ther 2002;24:616–28. Epub 2002/05/23.
18. Stewart WF, Van Rooyen JB, Cundiff GW, et al. Prevalence and burden of overactive bladder in the United States. World J Urol 2003;20:327–36. Epub 2003/06/18.
19. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993;31:247–63. Epub 1993/03/01.
20. Linacre JM. *WINSTEPS—Rasch analysis for all two-facet models*. Chicago, IL: MESA Press; 2000.
21. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978;43:561–73.
22. Wright BD, Linacre JM, Gustafson J-E, et al. Reasonable mean-square fit values. Rasch Meas Trans 1994;8:370.
23. Baker FB. editor. *The basics of item response theory. ERIC clearinghouse on assessment and evaluation*. College Park, MD; University of Maryland; 2001.
24. Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publications, Inc.; 1991.
25. Fayers PM, Hays R. *Assessing quality of life in clinical trials: Methods and practice*. USA: Oxford University Press; 2005.
26. Lai JS, Cella D, Kupst MJ, et al. Measuring fatigue for children with cancer: Development and validation of the pediatric Functional Assessment of Chronic Illness Therapy-Fatigue (pedsFACIT-F). J Pediatr Hematol Oncol 2007;29:471–9. Epub 2007/07/05.
27. Lai JS, Nowinski C, Victorson D, et al. Quality-of-life measures in children with neurological conditions: Pediatric Neuro-QOL. Neurorehabil Neural Repair 2012;26:36–47. Epub 2011/07/27.
28. Bjorner JB, Kosinski M, Ware JE Jr. Using item response theory to calibrate the Headache Impact Test (HIT) to the metric of traditional headache scales. Qual Life Res 2003;12:981–1002. Epub 2003/12/04.
29. Garratt AM. Rasch analysis of the Roland disability questionnaire. Spine (Phila Pa 1976) 2003;28:79–84. Epub 2003/01/25.
30. Tennant A, Hillman M, Fear J, et al. Are we making the most of the Stanford Health Assessment Questionnaire? Br J Rheumatol 1996;35:574–8. Epub 1996/06/01.
31. Coyne KS, Sexton CC, Irwin DE, et al. The impact of overactive bladder, incontinence and other lower urinary tract symptoms on quality of life, work productivity, sexuality and emotional well-being in men and women: Results from the EPIC study. BJU Int 2008;101:1388–95. Epub 2008/05/06.
32. Hu TW, Wagner TH. Health-related consequences of overactive bladder: An economic perspective. BJU Int 2005;96:43–5. Epub 2005/08/10.
33. Coyne KS, Zhou Z, Thompson C, et al. The impact on health-related quality of life of stress, urge and mixed urinary incontinence. BJU Int 2003;92:731–5. Epub 2003/11/18.
34. Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: A literature review. J Eval Clin Pract 1999;5:401–16. Epub 1999/12/01.
35. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: A structured review. J Eval Clin Pract 2006;12:559–68. Epub 2006/09/22.
36. Valderas JM, Kotzeva A, Espallargues M, et al. The impact of measuring patient-reported outcomes in clinical practice: A systematic review of the literature. Qual Life Res 2008;17:179–93. Epub 2008/01/05.
37. Hahn EA, Cella D, Bode RK, et al. Item banks and their potential applications to health status assessment in diverse populations. Med Care 2006;44:S189–97. Epub 2006/10/25.
38. Lai JS, Cella D, Dineen K, et al. An item bank was created to improve the measurement of cancer-related fatigue. J Clin Epidemiol 2005;58:190–7. Epub 2005/02/01.

Appendix A. OAB-q SF and Scoring Manual

OAB-q Short-Form

This questionnaire asks about how much you have been bothered by selected bladder symptoms during the past 4 weeks. Please place a ✓ or x in the box that best describes the extent to which you were bothered by each symptom during the past 4 weeks. There are no right or wrong answers. Please be sure to answer every question.

During the past 4 weeks, how bothered were you by.. .	Not at all	A little bit	Some-what	Quite a bit	A great deal	A very great deal
1. An uncomfortable urge to urinate?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
2. A sudden urge to urinate with little or no warning?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
3. Accidental loss of small amounts of urine?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
4. Nighttime urination?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
5. Waking up at night because you had to urinate?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
6. Urine loss associated with a strong desire to urinate?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

The previous questions asked about your feelings about individual bladder symptoms. For the following questions, please think about your overall bladder symptoms in the past 4 weeks and how these symptoms have affected your life. Please answer each question about how often you have felt this way to the best of your ability. Please place a ✓ or x in the box that best answers each question.

During the past 4 weeks, how often have your bladder symptoms	**None of the time**	**A little of the time**	**Some of the time**	**A good bit of the time**	**Most of the time**	**All of the time**
1. Caused you to plan "escape routes" to restrooms in public places?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
2. Made you feel like there is something wrong with you?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
3. Interfered with your ability to get a good night's rest?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
4. Made you frustrated or annoyed about the amount of time you spend in the restroom?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
5. Made you avoid activities away from restrooms (i.e., walks, running, hiking)?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
6. Awakened you during sleep?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
7. Caused you to decrease your physical activities (exercising, sports, etc.)?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
8. Caused you to have problems with your partner or spouse?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
9. Made you uncomfortable while traveling with others because of needing to stop for a restroom?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
10. Affected your relationships with family and friends?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
11. Interfered with getting the amount of sleep you needed?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
12. Caused you embarrassment?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
13. Caused you to locate the closest restroom as soon as you arrive at a place you have never been?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

To calculate a symptom bother score, create a summed score from the listed items and use the formula below the table to transform the value. This will provide symptom bother scores where higher score values are indicative of greater symptom bother and lower scores indicate minimal symptom bother.

Scale	Sum item values Part A	Lowest and highest possible raw scores	Possible raw score range
Symptom Bother	1–6	6, 36	30

Transformation for Symptom Severity raw scores ONLY:

$$\text{Transformed Score} = \frac{(\text{Actual raw score} - \text{lowest possible raw score})}{\text{Possible raw score range}} \times 100$$

For the HRQL subscales (coping, sleep, and social), create summed scores of the listed items for each individual subscale. Use the formula below the table to transform all values. Higher scores will be indicative of better HRQL.

Scale	Sum item values Part B	Lowest and highest possible raw scores	Possible raw score range
Total HRQL score	1–13	13, 78	65

Formula for transformation of HRQL raw scores:

$$\text{Transformed Score} = \frac{(\text{Highest possible score} - \text{Actual raw score})}{\text{Possible raw score range}} \times 100$$

Missing Items. For the subscale analyses, if $<$50% of the scale items are missing, the scale should be retained with the mean scale score of the items present used to impute a score for the missing items. If $\geq$50% of the items are missing, no scale score should be calculated, the subscale score should be considered missing.