Quality of Life Research (2005) 14: 849–855

The responsiveness of the Overactive Bladder Questionnaire (OAB-q)

Karin S. Coyne, Louis S. Matza & Christine L. Thompson
MEDTAP International, Inc., Center for Health Outcomes Research, 7101 Wisconsin Avenue, Suite 600, Bethesda, MD 20814, USA (E-mail: Coyne@medtap.com)

Accepted in revised form 15 June 2004

Abstract

Purpose: The purpose of this study was to examine the responsiveness of the Overactive Bladder questionnaire (OAB-q) during anti-muscarinic treatment. *Methods*: OAB patients were treated with tolterodine ER 4 mg/day for 12 weeks. The OAB-q and 3-day micturition diaries were collected at baseline, 4, and 12 weeks. The patients' and physicians' perceptions of treatment benefit were assessed at 4 and 12 weeks. Responsiveness of the OAB-q was examined with effect sizes and comparisons to other measures using ANOVAs, *t*-tests, and correlations. *Results*: A total of 865 patients completed the 12-week study (mean age 61 years; 73% female; 89% Caucasian). From baseline to 4 weeks, significant improvements ($p < 0.0001$) occurred in all OAB-q subscales, which were maintained through week 12. The OAB-q was highly responsive with subscale effect sizes ranging from 0.44 (social interaction) to 1.2 (symptom bother). Significant score changes in all OAB-q subscales ($p < 0.05$) were associated with reductions of $\geqslant$3 urgency episodes, $\geqslant$3 micturitions, or $\geqslant$1 incontinence episode per day. Improvements in OAB-q scales were associated with changes in patient and physician perceptions of treatment benefit. *Conclusions*: The OAB-q was highly responsive and demonstrated responsiveness to reductions in urinary urgency, frequency, and incontinence during antimuscarinic treatment of OAB. The OAB-q appears to be a useful outcome measure for treatments of OAB.

Key words: Health-related quality of life, Overactive bladder, Questionnaire, Responsiveness

Introduction

Overactive bladder (OAB) is characterized by symptoms of urinary urgency, with or without urge incontinence, and usually with increased urinary frequency and nocturia [1]. Community surveys in the United States and six European countries have found that symptoms of OAB are common among both men and women with overall prevalence estimates of greater than 16% among adults [2, 3]. Since OAB is primarily defined by symptoms rather than objective measures and has been shown to significantly impair health-related quality of life (HRQL) [4–6], evaluation of treatment effectiveness should include assessment of patient perceptions [7]. The Overactive Bladder Questionnaire (OAB-q) was developed to assess the patient's perceptions of symptom bother and impact on HRQL among patients with both continent and incontinent OAB [4].

The purpose of this study was to evaluate the responsiveness of the OAB-q in a clinical trial of treatment for OAB. The responsiveness of an instrument is the extent to which the instrument accurately reflects change in a patient's condition – that the instrument discriminates between patients who improve and those who do not [7–10]. An instrument without adequate responsiveness could fail to demonstrate statistical and clinical significance, regardless of the true treatment effect.

Methods

Data from a 12-week, multicenter, open-label clinical trial of the anti-muscarinic tolterodine ER

4 mg/day in OAB patients were used for this analysis. Details and results of this study have been published previously [11]. Briefly, inclusion criteria were: age $\geqslant$18; diagnoses of OAB with or without urge incontinence; and the ability to complete a micturition diary. Exclusion criteria were: predominant stress incontinence; an indwelling catheter or intermittent self-catheterization; symptomatic or recurrent urinary tract infection; hepatic or renal dysfunction; electrostimulation bladder training or pelvic floor exercises within 4 weeks of study entry; or pregnancy.

Measures

Overactive Bladder Questionnaire (OAB-q)

The OAB-q consists of an 8-item symptom bother scale and 25 HRQL items that form 4 subscales (coping, concern, sleep, and social interaction) and a total HRQL score. Patients rate each item on a 6-point Likert scale ranging from 'none of the time' to 'all of the time' for the HRQL items and 'not at all' to 'a very great deal' for the symptom bother items. Subscales are summed and transformed into scores ranging from 0 to 100; higher symptom bother scores indicate increasing symptom bother, higher HRQL scores indicate better HRQL.

Micturition diaries

Three-day micturition diaries assessed micturition frequency, incontinence episodes, and urgency episodes per 24 hours.

Assessment of response to treatment

At weeks 4 and 12, the physicians and patients separately rated the patients' response to treatment on a three-level scale (no benefit, a little benefit, and much benefit).

Statistical analysis

This secondary analysis included only the 'evaluable' patients with micturition diary data and at least one complete OAB-q subscale at baseline and 12 weeks. No data imputations for missing diaries or OAB-q subscale scores were performed. Change scores for the OAB-q subscales and micturition diaries were calculated from baseline to 12 weeks, baseline to 4 weeks, and 4–12 weeks with the primary analysis being baseline to 12-week change. All statistical tests were two-tailed and were conducted with Type I error probability of 0.05. Categorical variables are presented in terms of frequency and percents; means and standard deviations were calculated for continuous variables.

The responsiveness of the OAB-q was first examined using Spearman's correlations to assess the degree of association between OAB-q change scores and change in the micturition diaries. Correlation coefficients were interpreted as small (0.10), moderate (0.30), or large (0.50) following the guidelines proposed by Cohen [12]. For each micturition variable, categories of change from baseline were created (e.g., great, little, or no improvement) based upon data distribution and clinical judgment. Patients were also grouped according to their rating and the physician's rating of treatment benefit. OAB-q change scores were compared among the above-noted patient groups using analysis of variance models (ANOVAs) with Student–Newman–Keuls post hoc comparisons.

Effect sizes, a quantitative measure of change that standardizes the comparison between groups [13], were calculated for each OAB-q subscale by using the difference in mean score from baseline to week 12 and dividing by the standard deviation of baseline scores of all participants. Effect size was interpreted as small (0.20), moderate (0.50), or large (0.80) according to guidelines proposed by Cohen [12]. Additionally, the standardized response mean was calculated using the difference in mean score from baseline to week 12 divided by the standard deviation of this difference score. The responsiveness statistic was also calculated using the difference in mean score from baseline to week 12 divided by the standard deviation of the baseline scores of participants whose symptoms remained stable during treatment (i.e., no benefit group) [14].

Results

Demographic and clinical characteristics of the intent-to-treat population ($n = 1138$) were previously presented [11]. No statistically significant

Table 1. Baseline demographic and clinical characteristics

Patient characteristics	N = 865
Age mean years (SD)	61.0 (14.7)
Gender (n; % female)	636 (73.5%)
Race (n; %)	
Caucasian	772 (89.2%)
African-American	72 (8.3%)
Other	18 (2.1%)
Missing	3 (0.3%)
Symptom history (n; %)	
<6 months	49 (5.7%)
6 months–5 years	554 (64.1%)
>5 years	262 (30.3%)
Mean micturitions per 24 hours (SD)	13.3 (5.1)
Mean incontinence episodes per 24 hours (SD)	2.5 (3.5)
Mean urgency episodes per 24 hours (SD)	5.0 (4.2)

differences were present between the intent-to-treat and evaluable patient populations in baseline demographic or clinical characteristics (e.g., age, gender, duration of urinary symptoms, micturition diary data, or OAB-q scores). Of the 865 patients in the evaluable patient population, the mean age was 61 years, 73.5% were female, and 89.2% Caucasian (Table 1).

Baseline OAB-q scores indicate moderate symptom bother and HRQL impact with the sleep subscale affected greatest and the social interaction subscale affected the least. Statistically significant improvements were observed among all OAB-q subscales from baseline to 4 weeks (for all comparisons, $p < 0.0001$) with further improvements occurring at 12 weeks (all $p \leqslant 0.004$) (Figure 1). OAB-q subscale score changes from baseline to 12 weeks ranged from 10 (social interaction) to 26.1 (sleep) with the greatest improvements occurring in the first 4 weeks of treatment.

In addition to reductions in symptom bother and HRQL improvements, micturition diary variables also improved [11]. All correlations between OAB-q change scores and micturition diary changes from baseline to week 12 were statistically significant ($p < 0.001$) ranging from 0.13 to 0.35, which are considered small to moderate in magnitude. To further examine the effects of change in micturition variables and the OAB-q subscales, a series of ANOVAs compared OAB-q score changes by change in urinary urgency, frequency, and incontinence episodes. The OAB-q subscales were sensitive to change in all micturition diary variables; reductions in urinary urgency, frequency, and incontinence were associated with improvements in all OAB-q subscales. A reduction of three or more urgency episodes per day was associated with significant 15–36 point improvements in all OAB-q subscale scores (Table 2).

A reduction by three or more micturitions per day was associated with significant score changes of 10–28 points among all OAB-q subscales (Table 3). The coping and concern subscales had significant score changes (20-point change) with a reduction of one micturition/day. When analyses were

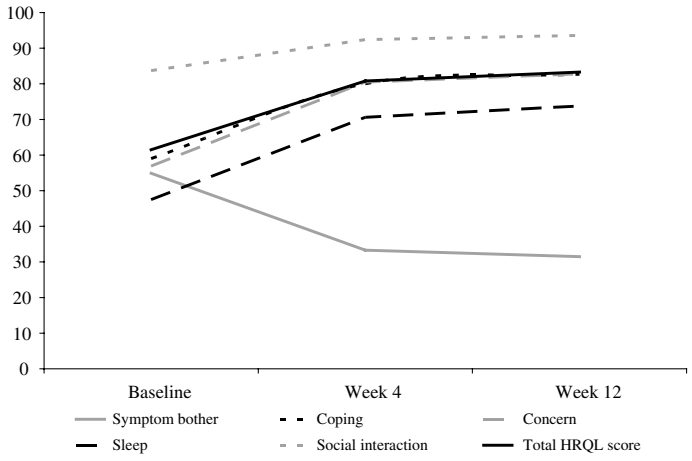

Figure 1. OAB-q scores at baseline, week 4, and week 12.

Table 2. Baseline to week 12 OAB-q Δ scores by change in urgency episodes from baseline to week 12

OAB-q subscales mean (SD)	Change in urgency episodes[a]			
	Little or no improvement (≤1) N = 225[d]	Slight improvement (>1–3) N = 280[e]	Moderate improvement (>3–4.5) N = 108[f]	Great improvement (>4.5) N = 226[g]
Δ Symptom bother	−15.7[a] (19.7)	−23.6[b] (19.8)	−31.4[c] (22.1)	−29.4[c] (21.5)
Δ Coping	17.4[a] (22.5)	20.5[a] (21.9)	32.8[b] (25.8)	31.2[b] (28.7)
Δ Concern	17.8[a] (22.7)	22.7[a] (22.5)	36.1[b] (27.8)	34.2[b] (27.8)
Δ Sleep	18.2[a] (27.2)	25.4[b] (28.8)	34.7[c] (29.8)	32.3[c] (29.6)
Δ Social interaction	7.1[a] (18.0)	7.1[a] (16.3)	14.8[b] (20.6)	14.6[b] (22.4)
Δ Total HRQL score	15.4[a] (18.4)	19.2[a] (17.9)	30.3[b] (21.4)	28.7[b] (22.8)

[a–c] Student–Newman–Keuls post hoc pairwise comparisons were performed. Superscript letters within each row indicate significant differences in OAB-q change scores among the four urgency groups at a level of $p < 0.05$. In general, greater reduction in urgency episodes was associated with significantly greater improvement in the OAB-q subscales.
[d] n varies from 196 to 225 depending on missing data for each subscale.
[e] n varies from 215 to 280 depending on missing data for each subscale.
[f] n varies from 92 to 108 depending on missing data for each subscale.
[g] n varies from 185 to 226 depending on missing data for each subscale.

conducted using percent change in micturition frequency, an improvement of 25% was associated with a significant score change (20-point plus change) in all subscales except social interaction. The findings for the change in number of incontinence episodes were similar to the findings for micturition frequency, with a reduction of more than one incontinence episode per day corresponding to dramatic improvements in all OAB-q subscales (Table 4). All score changes (range 13–34 point changes) were significant ($p < 0.05$) when compared to the group with no change in incontinence. A 50% reduction in incontinence episodes was associated with significant score changes (i.e., 25–30 points; $p < 0.05$) for all OAB-q scores except the social interaction subscale.

ANOVAs with post hoc pairwise comparisons of OAB-q change scores were performed using the physician-rated patient response to treatment (no change, little benefit, or much benefit) at 12 weeks (Figure 2). On all OAB-q scales except social interaction, the much benefit group demonstrated

Table 3. Baseline to week 12 OAB-q Δ scores by change in micturition frequency from baseline to week 12

OAB-q subscales mean (SD)	Change in micturition frequency[a]			
	Little or no improvement (≤1) N = 175[e]	Slight improvement (>1–3) N = 226[f]	Moderate improvement (>3–5) N = 207[g]	Great improvement (>5) N = 232[h]
Δ Symptom bother	−15.8[a] (21.4)	−19.3[a] (19.3)	−28.8[b] (19.5)	−32.7[b] (21.0)
Δ Coping	12.9[a] (19.5)	19.9[b] (22.0)	28.4[c] (25.6)	33.2[d] (27.8)
Δ Concern	16.5[a] (22.1)	21.2[a] (23.1)	28.9[b] (26.3)	35.9[c] (26.7)
Δ Sleep	13.0[a] (21.4)	20.9[b] (26.9)	29.0[c] (28.4)	39.9[d] (31.5)
Δ Social interaction	4.9[a] (14.5)	7.8[a] (17.5)	10.3[b] (18.5)	16.3[c] (23.4)
Δ Total HRQL score	12.3[a] (15.2)	17.8[b] (18.4)	24.8[c] (19.8)	31.8[d] (22.8)

[a–d] Student–Newman–Keuls post hoc pairwise comparisons were performed. Superscript letters within each row indicate significant differences in OAB-q change scores among the four micturition groups at a level of $p < 0.05$. In general, greater reduction in micturition frequency was associated with greater change in each OAB-q subscale.
[e] n varies from 163 to 175 depending on missing data for each subscale.
[f] n varies from 196 to 226 depending on missing data for each subscale.
[g] n varies from 166 to 207 depending on missing data for each subscale.
[h] n varies from 164 to 232 depending on missing data for each subscale.

Table 4. Baseline to week 12 OAB-q Δ scores by change in incontinence episodes from baseline to week 12

OAB-q subscales mean (SD)	Change in incontinence episodes[a]		
	Little or no improvement (≤0) N = 290[b]	Slight improvement (>0–1) N = 201[c]	Moderate improvement (>1) N = 343[d]
Δ Symptom bother	−18.9 (21.3)	−22.7 (20.9)	−28.7 (20.8)
Δ Coping	19.5 (23.1)	20.7 (23.8)	31.0 (26.9)
Δ Concern	19.8 (22.8)	22.3 (24.1)	33.9 (27.4)
Δ Sleep	23.5 (28.7)	24.9 (29.1)	29.9 (29.7)
Δ Social interaction	7.6 (17.1)	8.5 (18.7)	13.3 (21.4)
Δ Total HRQL score	17.9 (18.5)	19.1 (18.8)	27.7 (22.5)

[a] Student–Newman–Keuls post hoc pairwise comparisons were performed, using a significance level of $p < 0.05$. With regard to all OAB-q scales except sleep, the moderate improvement group demonstrated significantly greater change than the slight improvement group, and the Slight improvement group demonstrated significantly greater change than the little or no improvement group. On the sleep subscale, the moderate improvement group demonstrated significantly greater change than the little or no improvement group, but the slight improvement group was not significantly different from either of the other two groups. In general, greater reduction in the number of incontinence episodes was associated with greater improvements in each OAB-q subscale score.
[b] n varies from 232 to 290 depending on missing data for each subscale.
[c] n varies from 167 to 201 depending on missing data for each subscale.
[d] n varies from 285 to 343 depending on missing data for each subscale.

significantly greater change than the little benefit group and the little benefit group demonstrated significantly greater change than the no change group. On the social interaction scale, the much benefit and little benefit groups demonstrated significantly greater change than the no change group, but were not significantly different from each other. Similar results were found with the patient perception of treatment benefit analysis.

All of the OAB-q subscales demonstrated moderate to large effect sizes from baseline to 12 weeks ranging from 0.44 for the social interaction subscale to −1.23 for the symptom bother subscale (Table 5). From baseline to week 4, the effect sizes were similar but of slightly lesser magnitude,

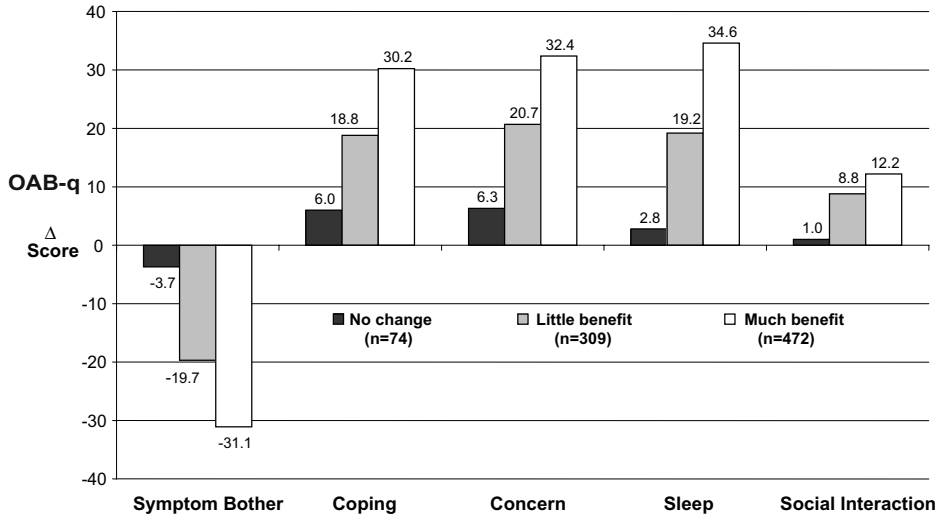

Figure 2. OAB-q change scores by physician's assessment of treatment benefit.

Table 5. Effect size, standardized response mean, and responsiveness statistic for OAB-q scales: baseline to week 12

OAB-q subscales	Effect size	Response mean	Responsiveness statistic
Symptom bother (n = 705)*	−1.23	−1.11	−1.10
Coping (n = 846)	0.84	0.94	0.79
Concern (n = 855)	0.95	1.00	0.89
Sleep (n = 855)	0.85	0.89	0.79
Social interaction (n = 830)	0.44	0.52	0.40
Total HRQL score (n = 810)	0.97	1.05	0.89

*Sample size varies among subscales due to missing data.

ranging from 0.38 for social interaction to 1.16 for symptom bother. Effect sizes from week 4 to week 12 were of a small magnitude (0.07–0.13) reflecting the relatively small change that occurred between these two visits. The standardized response mean and the responsiveness statistic followed the same pattern as the effect sizes.

Discussion

The OAB-q, with its demonstrated reliability and validity [4], is highly responsive to treatment-related change. The OAB-q was able to discriminate among levels of change in all micturition diary variables, which are frequent clinical outcomes for OAB treatments, as well as physician and patient ratings of treatment benefit. Reductions of three or more urgency episodes or micturitions per day were associated with significant score changes in all OAB-q subscales indicating that this level of clinical improvement translated into significant HRQL benefits and reductions in bothersome symptoms. Importantly, for the symptom bother, sleep, and concern subscales, changes of lesser magnitude in urinary urgency and frequency were associated with significant score changes suggesting that patients perceive an HRQL improvement with lesser clinical change.

While the symptom bother subscale clearly was the most sensitive to change, the concern, sleep, and coping subscales were also highly responsive. The social interaction subscale was the least responsive to change, possibly due to longer-term patient adaptation. It may be that patients with chronic urinary problems adjust their lifestyle to accommodate their illness, thus reducing their social activities. If so, a longer treatment follow-up period would be necessary to detect changes in social interaction.

The usefulness of a patient-reported outcome measure depends not only on its reliability and validity, but also on its ability to detect treatment-related change over time [15]. Confidence in findings depends upon the amassed evidence from different methods and studies. Greater confidence can be placed in the interpretation of clinically significant effects with consistent evidence from multiple sources. In the current study, the OAB-q demonstrated responsiveness in a range of different analyses, including large to moderate effect sizes as well as significant discriminative ability with clinical measures, patient perceptions, and physician ratings. The consistency of our findings with varying clinical outcomes provides assurance that the observed score changes are valid and clinically meaningful.

One limitation of this study is that the data are from a clinical study that was not placebo-controlled. Although the OAB-q was found to be responsive to perceived symptom change and micturition diary data, some of this perceived change is likely to be a result of placebo effects as has been noted in previous research [16]. Thus, in this study it is not possible to determine the responsiveness of the OAB-q due to true treatment effects, as opposed to placebo effects. Future research estimating responsiveness with placebo-controlled data is necessary to address this issue.

Patient-reported outcomes including symptom bother and HRQL are critical for providing a thorough understanding of treatment effects, particularly for symptom-defined conditions such as OAB. For OAB, patient-reported outcomes provide added value and may actually be more important than clinical outcomes in understanding the overall effects of treatments on patients' lives. The current findings suggest that the OAB-q is highly responsive and able to discriminate between

OAB patients who improve and those who do not. Consequently, the OAB-q appears to be a useful outcome measure for evaluating treatment-related change in symptom bother and HRQL in clinical trials and practice with OAB patients.

Acknowledgements

This research was funded by Pfizer, Inc. The authors would like to thank Jodi Shorr and Fritz Hamme for their production assistance.

References

1. Abrams P, Cardozo L, Fall M, et al. The standardisation of terminology of lower urinary tract function: Report from the Standardisation Sub-committee of the International Continence Society. Am J Obstet Gynecol 2002; 187(1): 116–126.
2. Milsom I, Abrams P, Cardozo L, et al. How widespread are the symptoms of an overactive bladder and how are they managed? A population-based prevalence study. BJU Int 2001; 87(9): 760–766.
3. Stewart WF, Van Rooyen JB, Cundiff GW, et al. Prevalence and burden of overactive bladder in the United States. World J Urol 2003; 20(6): 327–336.
4. 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(6): 563–574.
5. Jackson S. The patient with an overactive bladder – symptoms and quality-of-life issues. Urology 1997; 50(Suppl. 6A): 18–22; discussion 23–14.
6. Liberman JN, Hunt TL, Stewart WF, et al. Health-related quality of life among adults with symptoms of overactive bladder: Results from a U.S. community-based survey. Urology 2001; 57(6): 1044–1050.
7. Leidy NK, Revicki DA, Geneste B. Recommendations for evaluating the validity of quality of life claims for labeling and promotion. Value in Health 1999; 2(2): 113–127.
8. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance. J Chronic Dis 1986; 39(11): 897–906.
9. Hays RD, Hadorn D. Responsiveness to change: An aspect of validity, not a separate dimension. Qual Life Res 1992; 1(1): 73–75.
10. Revicki DA, Osoba D, Fairclough D, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. Qual Life Res 2000; 9(8): 887–900.
11. 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(4): 616–628.
12. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
13. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989; 27(Suppl. 3): S178–S189.
14. Sprangers MA, Moinpour CM, Moynihan TJ, et al. Assessing meaningful change in quality of life over time: A users' guide for clinicians. Mayo Clin Proc 2002; 77(6): 561–571.
15. Guyatt G, Walter S, Norman G. Measuring change over time: Assessing the usefulness of evaluative instruments. J Chronic Dis 1987; 40(2): 171–178.
16. Kelleher CJ, Reese PR, Pleil AM, et al. Health-related quality of life of patients receiving extended-release tolterodine for overactive bladder. Am J Manag Care 2002; 8(Suppl. 19): S608–S615.

Address for correspondence: Karin S. Coyne, MEDTAP International, Inc., Center for Health Outcomes Research, 7101 Wisconsin Avenue, Suite 600, Bethesda, MD 20814, USA
E-mail: Coyne@medtap.com