Volume 8 • Supplement 1 • 2005
VALUE IN HEALTH

# The Validation of Patient-Rated Global Assessments of Treatment Benefit, Satisfaction, and Willingness to Continue—The BSW

Andreas M. Pleil, PhD,[1] Karen S. Coyne, PhD,[2] Pat Ray Reese, PhD,[3] Zhanna Jumadilova, MD,[1] Eric S. Rovner, MD,[4] Con J. Kelleher, MD[5]

[1]Pfizer Inc, San Diego, CA, USA; [2]The MEDTAP Institute at UBC, Bethesda, MD, USA; [3]Reese Associates Consulting, LLC, Cary, NC, USA; [4]MUSC, Charleston, SC, USA; [5]Guy's and St. Thomas' NHS Foundation Trust, London, UK

## ABSTRACT

**Objective:** This study evaluated the validity of three single-item, patient-rated, interviewer-administered, global assessments of treatment benefit, satisfaction with treatment and willingness to continue treatment, collectively referred to as the BSW.

**Methods:** The BSW, micturition diaries, the Overactive Bladder Questionnaire (OAB-q) and the King's Health Questionnaire (KHQ) were included in part or in total in three OAB clinical trials. Discriminant validity for full and dichotomized responses was assessed with ANOVAs models and correlations were used to evaluate construct validity.

**Results:** The BSW demonstrated significant differences among the majority of the response levels on all measures of micturitions in all studies. The BSW also demonstrated discriminant validity with the OAB-q and the KHQ. BSW measures demonstrated significant differences among the change scores for all subscales of the OAB-q and the majority of the KHQ domains with both full and dichotomized responses. Patients who were dissatisfied with treatment and those unwilling to continue treatment also reported significantly worse OAB-q and KHQ scores compared with those who were satisfied with treatment or willing to continue treatment. BSW measures were moderately correlated with the micturition variables, moderate to strongly correlated with the OAB-q and weak to moderately with the KHQ, providing support for the construct validity of the BSW measures.

**Conclusions:** The BSW is a useful tool to capture patients' global impressions of three key elements of treatment outcome: a perceived benefit, satisfaction with treatment, and the willingness to continue treatment, and can facilitate patient–physician communication as well as be informative to researchers.

*Keywords:* incontinence, overactive bladder, quality of life, satisfaction, treatment benefit.

## Introduction

Over the past decade, patient-reported outcomes (PRO) have increasingly become accepted as important measures of clinical outcome, particularly when assessing symptom-based conditions such as gastroesophageal reflux disease (GERD) [1], migraine [2], or overactive bladder (OAB) [3,4]. The symptom complex of OAB, which includes urinary urgency, with or without urge incontinence, and usually with increased urinary frequency and nocturia [5], has been shown to cause significant symptom bother and to reduce health-related quality of life (HRQL) [2,6,7]. It is this symptom bother and HRQL impairment that most commonly causes patients with OAB to seek treatment. Thus, when assessing treatment outcome, the patient's subjective opinion is of paramount importance. Using PROs when assessing OAB treatment enhances the clinical evaluation of the treatment by assessing changes in symptom bother and improvement or declines in HRQL [8].

PROs such as HRQL and treatment satisfaction have been assessed using multi-item and single-item measures. Multi-item measures are a rich source of information regarding the many facets of patient's lives that are affected by a condition (e.g., disease impact on HRQL includes such domains as sleep, physical and emotional function and social activi-

*Address correspondence to:* Andreas Pleil, Director/Site Leader Worldwide Outcomes Research, Pfizer Inc, Building 95, 11095 Torreyana Road, San Diego, CA, USA 92121. E-mail: andreas.m.pleil@pfizer.com

doi:10.1111/j.1524-4733.2005.00069.x

 1098-3015/05/S25 S25–S34

ties). Generally, these measures include the patient's perspective and are designed on the basis of a sound understanding of the ways a condition may affect patients' lives.

Nevertheless, single-item measures, such as both physician- and patient-rated global assessments are used routinely as an indicator of treatment outcomes [9–12]. To provide a global assessment, physicians' evaluate multiple clinical factors, comparing these with their extensive clinical experience, and provide an informed judgment. Typically, patient-rated global assessments receive less consideration; however, their usefulness should not be underestimated. Patient-rated assessments reflect the individual patient's perspective by integrating the various aspects of importance to them and providing a single outcome response. The aspects considered in the patient's perspective may differ from patient to patient, as it reflects the individual's needs, concerns, and values. Thus, the underlying assumption of patient-rated global assessments is that the patient will weigh all factors related to a condition/disease (e.g., risk/benefit of a treatment; consideration of multiple symptoms rather than one symptom) and provide a response that reflects their perspective of the construct being measured, just as a physician would when providing a clinical global assessment.

Patient-rated, global assessments of treatment benefit, satisfaction with the treatment and willingness to continue to use a treatment are often considered in determinations of treatment effectiveness. Patient perceptions of treatment benefit are assumed to include an evaluation of the balance between the experience of treatment risks and benefits. Satisfaction with treatment is thought to include an evaluation of the individual's needs, perceived benefit, concerns and expectations, thus converting a complex construct into a global response. Willingness to continue treatment, generally considered an indicator of future adherence, can be linked to satisfaction and perceived benefit. Asking patients questions regarding treatment benefit, satisfaction, and willingness to continue facilitates patient–physician communication and builds consensus in treatment decisions. Nevertheless, many physicians have had less confidence in these patient-rated assessments because the validity of these measures is rarely established.

The purpose of this research was to retrospectively evaluate the validity of three, single-item, global assessments: collectively referred to as the benefit, satisfaction, and willingness to continue (BSW) measure. This is important because a validated, structured BSW measure can facilitate physician efforts to gain an understanding of how patients value their treatment. Additionally, the BSW measure can provide a quick and convenient way to determine if there are patient issues related to their treatment that require additional exploration.

## Methods

The BSW was used in part or in total in three recently completed large-scale, 12-week, randomized, double-blind, placebo-controlled clinical trials of tolterodine in the treatment of patients with OAB (Table 1). All patients included in the studies fulfilled the ICS criteria for a diagnosis of OAB syndrome including the presence of urinary frequency of ≥8 micturitions/24 h and urgency with or without urge incontinence. Patients participating in Studies 1 and 2 had an additional requirement of nocturia with an average of 2.5 or more nighttime voids over a 7-day period. Patients from Study 3 had an additional requirement of urgency incontinence with five or more episodes in a 7-day period. Micturition characteristics were collected using a patient diary. All patients were 18 years of age or older and recruited without regard to sex. The studies were conducted in accordance with the Declara-

**Table 1** Summary of tolterodine studies

| | Study 1 | Study 2 | Study 3 |
|---|---|---|---|
| Sample size | 596 | 555 | 1177 |
| Location | USA and Chile | International | International |
| 7-day micturition diary | Frequency, urgency rating, nocturia | Frequency, urgency rating, nocturia | Frequency, incontinence episodes |
| Benefit | ✓ | ✓ | ✓ |
| Satisfaction | ✓ | ✓ | NA |
| Willingness to continue | ✓ | ✓ | NA |
| OAB-q | ✓ (USA only) | ✓ (Canada and USA only, n = 85) | NA |
| KHQ | ✓ (Chile only) | ✓ | ✓ |

NA, not assessed.

tion of Helsinki, local independent ethics committee/institutional review board requirements, and good clinical practice guidelines. Patients provided written informed consent before study entry.

### *Micturition Diary*

A printed micturition diary was used to collect data for the clinical efficacy assessments in all three trials, although the structure varied slightly due to differences in study design and end points. For all three studies, patients recorded every micturition episode. For Studies 1 and 2, patients recorded the time of each micturition so that daytime and nighttime voids could be discerned. Additionally, in Studies 1 and 2, patients rated their level of urgency associated with each micturition as follows:

1. No urgency: I felt no need to empty my bladder but did so for other reasons.
2. Mild urgency: I could postpone voiding as long as necessary without fear of wetting myself.
3. Moderate urgency: I could postpone voiding for a short time without fear of wetting myself.
4. Severe urgency: I could not postpone voiding but had to rush to the toilet in order not to wet myself.
5. Urge incontinence: I leaked before arriving at the toilet.

To account for both urinary frequency and urgency severity ratings, an urgency-frequency severity rating was calculated by summing each urgency severity rating per day, which implicitly reflects each patient's urinary frequency. This rating was analyzed as a continuous variable. Urgency level was not assessed in Study 3.

### *Patient-Reported Outcome Measures*

*The benefit, satisfaction, and willingness to continue measure.* The BSW consists of three, single-item measures designed to capture the patient's perception of the effect of treatment in terms of the relative benefit, their satisfaction, and their intention or willingness to continue on therapy (Appendix A). The Perception of Treatment Benefit question asks patients if they perceived a benefit from treatment. If the patient responds yes, the patient is then asked if the perceived benefit was of little benefit or much benefit. The Satisfaction question is similar as patients are asked if they are satisfied with treatment. If the response is yes, the patient is then asked if they are a little satisfied or very satisfied. If the response is no, the patient is asked if they are a little dissatisfied or very dissatisfied. The Willingness to Continue question was formatted in the same manner as the Satisfaction question with the responses being a little bit willing or very willing for yes responses and a little unwilling or very unwilling for no responses. A neutral mid-point was omitted to avoid ambiguous responses. The BSW was administered by the investigator or designated site personnel in the local language as a standardized interview during the follow-up visits. The BSW can potentially be self-administered; however, this method of administration has not been tested. The PRO measures were translated and validated according to generally accepted translation guidelines [13] as necessary.

*Overactive bladder questionnaire (OAB-q).* The OAB-q is a 33-item, disease-specific measure designed to assess symptom bother and the impact of OAB symptoms on HRQL [2]. The OAB-q consists of a symptom bother scale and four HRQL domains: Concern, Coping, Sleep, and Social interactions. The OAB-q is a reliable and valid instrument that can discriminate between normal and clinically diagnosed continent and incontinent patients with OAB. The scores are transformed into a 0–100 scale. The Symptom Bother and HRQL subscale scores are inversed with high scores on the Symptom Bother scale indicating increased symptom severity and high scores on the HRQL subscales indicating better HRQL. The OAB-q has been shown to be highly responsive to treatment [8].

*King's health questionnaire (KHQ).* The KHQ is a 33-item, multidimensional, disease specific, widely used measure of HRQL and symptom severity in patients with OAB and incontinence. Developed by Kelleher and associates, the KHQ is the result of several years of refinement in more than 1000 patients referred to a tertiary urogynecology unit [6,14]. The KHQ consists of the following summated, multi-item HRQL domains: Role Limitations, Physical Limitations, Social Limitations, Personal Relationships, Emotions, Sleep and Energy, and Severity (coping) Measures. In addition, two one-item questions address Incontinence Impact and General Health Perceptions. The KHQ also includes the multi-item Symptom Severity scale that measures the severity of urinary symptoms. The KHQ is a reliable and valid instrument that can discriminate between normal and clinically diagnosed OAB patients [14,15]. The KHQ domains are scored on a 0 (best) to 100 (worst) scale.

### *Statistical Analysis*

This validation analysis includes patients with both baseline and follow-up clinical and PRO outcomes, excluding patients with major protocol violations, as decided before the randomization code was broken, and patients withdrawn from the study. ANOVAs (PROC GLM) were used to assess the ability of the BSW scale to discriminate among changes in the micturition diary variables, the OAB-q, and the KHQ scores over 12 weeks. All models controlled for age and sex. Scheffe post hoc pairwise comparisons were performed with *P*-values adjusted for multiple comparisons. To examine the discrimination of responses, the Satisfaction and Willingness to Continue questions were analyzed both by full response options (i.e., four groups) and by collapsing responses into dichotomous categories (i.e., satisfied/dissatisfied and willing/unwilling). Spearman's correlation coefficients with the micturition variables, the OAB-q, and the KHQ domains were also used to evaluate construct validity of the BSW.

## Results

The baseline demographics of the three studies varied with what appears to be a slightly younger cohort in Study 1 and more male patients in Studies 1 and 2 than in Study 3 (48.7%, 39.6%, and 20.2%, respectively) (Table 2). As expected based on inclusion criteria, Study 3 patients reported a greater number of incontinence episodes than Study 1 and 2 patients although statistical tests were not performed.

### *Change in Micturitions*

Table 3 shows that each of the BSW measures demonstrated statistically significant differences among the majority of the response levels on all measures of micturitions in all studies. Patients who reported "no treatment benefit" experienced a reduction from baseline of less than one micturition per 24 h for each of the studies compared with almost two micturitions for those who reported "little benefit" and approximately three micturitions for those perceiving "much benefit" from their treatment. Additionally, those who were dissatisfied with treatment also experienced a smaller reduction in micturition frequency as did those unwilling to continue treatment compared with those who were satisfied and those willing to continue treatment.

Other micturition measures followed a similar pattern. The mean changes in the number of nighttime voids were consistent across the two nocturia studies: Study 1: –0.39 and Study 2: –0.30 for those perceiving "no benefit," Study 1: –0.80, and Study 2: –0.79 for those reporting "little benefit," and Study 1: –1.35 and Study 2: –1.41 for those reporting "much benefit" (Table 3). Similarly, those who were dissatisfied with treatment also experienced a smaller reduction in nighttime voids as did those unwilling to continue treatment compared with those who were satisfied and those willing to continue treatment. Table 3 also shows that patients in Study 3 reported significantly fewer reductions in incontinence episodes for those perceiving "no benefit" compared with those reporting "little benefit" or "much benefit."

### *Overactive Bladder Questionnaire*

Each of the BSW measures demonstrated discriminant validity based on the OAB-q results for Study 1 (Table 4). In this study, each of the BSW measures demonstrated statistically significant differences among the 12-week change scores for all subscales of the OAB-q. For treatment benefit, all OAB-q change from baseline scores were significantly improved for those who perceived "much benefit" compared with those who perceived "little" or "no benefit." The largest differences were observed in the Sleep and Symptom Bother domains where the

**Table 2** Baseline demographics of three studies

| | Study 1 N = 596 | Study 2 N = 555 | Study 3 N = 1177 |
|---|---|---|---|
| Age in years (SD) | 58.4 (13.2) | 60.6 (13.5) | 61.0 (14.0) |
| Sex, n female (%) | 306 (51.3) | 335 (60.4) | 939 (79.8) |
| Race | | | |
| White, n (%) | 511 (85.7) | 508 (91.5) | 1108 (94.1) |
| African American, n (%) | 48 (8.1) | 5 (0.9) | 47 (4.0) |
| Other, n (%) | 12 (2.0) | 10 (1.8) | 21 (1.9) |
| Missing, n (%) | 25 (4.2) | 32 (5.8) | 1 (0.1) |
| Micturition frequency per 24/h (SD) | 13.7 (3.3) | 13.7 (4.2) | 11.1 (3.8) |
| Incontinence episodes per 24/h (SD) | 0.5 (1.2) | 0.3 (0.9) | 3.3 (3.1) |
| Nocturia episodes per night (SD) | 3.5 (0.9) | 3.7 (1.3) | NA |
| Sum of urgency severity (SD) | 269.8 (101) | 265.6 (106) | NA |

**Table 3** Change in micturition variables by BSW at week 12 controlling for age and sex

| Micturition variables LS mean (SE) | Patient perception of treatment benefit† | | | Patient satisfaction with treatment‡ | | Patient willingness to continue with treatment‡ | |
|---|---|---|---|---|---|---|---|
| | No benefit | Little benefit | Much benefit | Dissatisfied | Satisfied | Unwilling | Willing |
| Study 1 | (N = 218) | (N = 166) | (N = 157) | (N = 200) | (N = 335) | (N = 209) | (N = 332) |
| Mean number of micturitions per 24 h | −0.89 (0.18) | −2.08 (0.20) | −3.36 (0.21)*** | −0.98 (0.19) | −2.51 (0.15)*** | −0.78 (0.18) | −2.67 (0.14)*** |
| Mean number of nocturia episodes per night | −0.39 (0.06) | −0.80 (0.07) | −1.35 (0.08)*** | −0.41 (0.07) | −1.00 (0.05)** | −0.41 (0.07) | −1.01 (0.05)** |
| Sum of urgency severity | −12.4 (5.26) | −47.9 (6.04) | −82.1 (6.25)*** | −13.76 (5.47) | −59.4 (4.23)*** | −11.9 (5.33) | −61.7 (4.23)*** |
| Study 2 | (N = 192) | (N = 175) | (N = 167) | (N = 190) | (N = 326) | (N = 196) | (N = 320) |
| Mean number of micturitions per 24 h | −0.82 (0.20) | −1.76 (0.21) | −3.49 (0.21)** | −0.79 (0.20) | −2.66 (0.16)*** | −0.80 (0.20) | −2.67 (0.16)*** |
| Mean number of nocturia episodes per night | −0.30 (0.08) | −0.79 (0.08) | −1.41 (0.08)*** | −0.30 (0.08) | −1.12 (0.06)*** | −0.34 (0.08) | −1.11 (0.06)*** |
| Sum of urgency severity | −14.0 (5.54) | −37.8 (5.81) | −83.8 (5.95)* | −14.3 (5.78) | −60.1 (4.40)*** | −14.3 (5.69) | −61.0 (4.44)*** |
| Study 3 | (N = 370) | (N = 365) | (N = 440) | | | | |
| Mean number of micturitions per 24 h | −0.75 (0.16) | −1.81 (0.16) | −2.57 (0.15)*** | | | | |
| Mean number of incontinence episodes per 24 h | −0.83 (0.12) | −1.51 (0.12) | −2.21 (0.11)*** | | | | |

†Pairwise comparison between means was performed using Scheffe's test adjusting for multiple comparisons.
‡*t*-test comparison between dichotomous groups.
*$P$ <0.05, **$P$ <0.01, ***$P$ <0.001.

absolute differences in change scores between "much benefit" and "no benefit" groups were 37.7 and 25.3 points, respectively. In Study 2, where only 85 patients provided responses to the OAB-q, the absolute differences between Sleep and Symptom bother change scores were similar (34.7 and 22.7 points, respectively). With the exception of the Social Interaction subscale, most of the comparisons between benefit levels were statistically significant, although there were fewer statistically significant differences between "no benefit" and "little benefit" responses in this study than observed in the larger Study 1.

Patients who were dissatisfied with treatment also reported significantly lower OAB-q subscale change scores compared with those who were satisfied with treatment (Table 4). The largest differences were seen in the Sleep and Symptom Bother subscale change scores; absolute differences of 23.2 and 16.3 points, respectively. The finding of significant differences among response levels was true whether satisfaction was dichotomized as dissatisfied or satisfied, as in Table 4, or analyzed as the original four levels of satisfaction, as in Table 5. With the exception of the Social Interaction subscale, the results from the Study 2 were similar, although fewer comparisons between the levels of dissatisfaction were statistically significant.

Findings for the willingness to continue treatment measure were similar to the other BSW measures in that patients who were willing to continue treatment reported significantly better OAB-q scale score changes on all subscales (Table 4). Similarly, the largest differences were seen in the Sleep, Symptom Bother, and Concern subscales; absolute differences of 22.4, 15.7, and 14.6 points, respectively. For Study 2, all OAB-q subscale change scores, with the exception of Social Interaction, were significantly higher for those willing to continue treatment compared with those unwilling to do so.

### *King's Health Questionnaire*

Each of the BSW measures demonstrated discriminant validity using the KHQ domains, with the exception of General Health Perceptions and Personal Relationships in Study 3 (Tables 6 and 7). Most KHQ domain changes from baseline to week 12 scores were significantly greater for those who perceived higher levels of benefit compared with those who perceived lower levels of benefit (Table 6). The largest differences in change scores were seen in the Incontinence Impact, Role Limitations, and Physical Limitations domains (absolute

**Table 4** Study 1: OAB-q change scores by perception of treatment benefit at 12 weeks controlling for age and sex

| OAB-q subscales LS mean (SE) | Patient perception of treatment benefit† | | | Patient satisfaction with treatment‡ | | Patient willingness to continue with treatment‡ | |
|---|---|---|---|---|---|---|---|
| | No benefit (N = 216) | Little benefit (N = 165) | Much benefit (N = 157) | Dissatisfied (N = 199) | Satisfied (N = 333) | Unwilling (N = 207) | Willing (N = 331) |
| Symptom bother | –4.1 (1.1) | –16.3 (1.3) | –29.4 (1.3)*** | –5.3 (1.3) | –21.6 (1.0)*** | –5.7 (1.3) | –21.4 (1.0)*** |
| Coping | 3.6 (1.3) | 10.7 (1.5) | 23.9 (1.6)** | 4.5 (1.4) | 16.1 (1.1)*** | 5.2 (1.4) | 15.7 (1.1)*** |
| Concern | 0.9 (1.3) | 12.1 (1.4) | 24.0 (1.5)*** | 1.9 (1.4) | 16.5 (1.0)*** | 2.0 (1.3) | 16.6 (1.0)*** |
| Sleep | 2.5 (1.5) | 19.4 (1.7) | 40.2 (1.8)*** | 4.5 (1.7) | 27.7 (1.3)*** | 5.0 (1.7) | 27.4 (1.3)*** |
| Social interaction | 0.5 (1.0) | 4.7 (1.2) | 10.7 (1.2)* | 0.0 (1.0) | 7.2 (0.8)** | 0.8 (1.0) | 6.8 (0.8)*** |
| HRQL total scale | 1.9 (1.1) | 11.7 (1.2) | 24.5 (1.3)*** | 2.9 (1.2) | 16.8 (0.9)*** | 3.4 (1.2) | 16.6 (0.9)*** |

†Pairwise comparison between means was performed using Scheffe's test adjusting for multiple comparisons. All pairwise comparisons were significant *P*-values.
‡*t*-test comparison between dichotomous groups.
**P* <0.05., ***P* <0.01, ****P* <0.001.

differences between "much benefit" and "no benefit" groups of 23.6, 23.2, and 23.2 points, respectively; Table 6). For Study 2, 21 of the 27 comparisons, nine scales with three benefit comparisons, were statistically significant, with the comparison between "no benefit" versus "little benefit" less frequently significant than observed in Study 3. The largest differences in change scores were seen in the Sleep/Energy and Incontinence Impact domains, absolute differences between "much benefit" and "no benefit" groups of 30.7 and 24.6 points, respectively, in Study 2.

For the satisfaction with treatment measure, Table 7 shows that all KHQ domain change scores were greater for those who were satisfied compared with those who were dissatisfied in Study 2. The largest differences were seen in the Sleep/Energy, Physical Limitation, and Incontinence Impact domains; absolute differences of 20.9, 15.6, and 15.4 points, respectively. Similarly, all KHQ domain

**Table 5** Study 1: OAB-q change scores by satisfaction with treatment at 12 weeks controlling for age and sex

| OAB-q subscales LS mean (SE) | Patient satisfaction with treatment | | | | Overall *F*-value | *P*-values for pairwise comparisons† |
|---|---|---|---|---|---|---|
| | Very dissatisfied (N = 93) | A little dissatisfied (N = 94) | A little satisfied (N = 131) | Very satisfied (N = 200) | | |
| Symptom bother | –3.2 (1.8) | –7.0 (1.8) | –15.9 (1.5) | –25.6 (1.2) | 31.2*** | 2***, 3***, 4**, 5***, 6*** |
| Coping | 2.0 (2.1) | 7.0 (2.0) | 10.2 (1.7) | 19.9 (1.4) | 17.0*** | 2*, 3***, 5***, 6*** |
| Concern | –2.3 (1.9) | 6.1 (1.9) | 10.4 (1.6) | 20.6 (1.3) | 28.3*** | 1*, 2***, 3***, 5***, 6*** |
| Sleep | 0.3 (2.4) | 8.5 (2.4) | 16.6 (2.0) | 35.1 (1.6) | 43.6*** | 2***, 3***, 5***, 6*** |
| Social interaction | –1.5 (1.5) | 1.3 (1.5) | 3.9 (1.3) | 9.2 (1.0) | 11.0*** | 3***, 5***, 6* |
| HRQL total scale | –0.3 (1.7) | 6.0 (1.6) | 10.3 (1.4) | 21.0 (1.1) | 34.3*** | 2***, 3***, 5***, 6*** |

†Pairwise comparison between means was performed using Scheffe's test adjusting for multiple comparisons. 1 = very dissatisfied versus a little dissatisfied, 2 = very dissatisfied versus a little satisfied, 3 = very dissatisfied versus very satisfied, 4 = a little dissatisfied versus a little satisfied, 5 = a little dissatisfied versus very satisfied, and 6 = a little satisfied versus very satisfied.
**P* <0.05, ***P* <0.01, ****P* <0.001.

**Table 6** Study 3: KHQ change scores by perception of treatment benefit at week 12 controlling for age and sex

| KHQ subscales LS mean (SE) | Patient perception of treatment benefit | | | *P*-values† |
|---|---|---|---|---|
| | No benefit (N = 370) | Little benefit (N = 365) | Much benefit (N = 440) | |
| General health perception | 1.7 (0.9) | –0.8 (0.9) | –2.2 (0.8) | 2** |
| Incontinence impact | –3.6 (1.4) | –13.0 (1.4) | –27.2 (1.3) | 1***, 2***, 3*** |
| Role limitations | –5.2 (1.5) | –12.8 (1.5) | –28.4 (1.4) | 1**, 2***, 3*** |
| Physical limitations | –2.0 (1.5) | –12.2 (1.5) | –25.2 (1.4) | 1***, 2***, 3*** |
| Social limitations | –0.6 (1.2) | –7.4 (1.2) | –15.1 (1.1) | 1***, 2***, 3*** |
| Personal relationships | –1.7 (1.7) | –3.7 (1.8) | –10.5 (1.6) | 2**, 3* |
| Emotions | –2.3 (1.3) | –8.4 (1.3) | –17.4 (1.2) | 1**, 2***, 3*** |
| Sleep/energy | –2.3 (1.2) | –8.4 (1.2) | –14.1 (1.1) | 1**, 2***, 3** |
| Severity measures | –2.3 (1.1) | –7.7 (1.1) | –19.6 (1.0) | 1, 2***, 3*** |

†Pairwise comparison between means was performed using Scheffe's test adjusting for multiple comparisons. 1 = No benefit versus little benefit, 2 = No benefit versus much benefit, and 3 = little benefit versus much benefit.
**P* <0.05, ***P* <0.01, ****P* <0.001.

**Table 7** Study 2: KHQ change scores by dichotomous satisfaction with treatment at 12 weeks controlling for age and sex

| | Patient satisfaction with treatment | | | Patient willingness to continue with treatment | | |
|---|---|---|---|---|---|---|
| KHQ subscales LS mean (SE) | Dissatisfied (N = 187) | Satisfied (N = 322) | *P*-values† | Unwilling (N = 193) | Willing (N = 315) | *P*-values |
| General health perception | –0.4 (1.4) | –4.0 (1.0) | 0.04 | –1.7 (1.4) | –3.5 (1.1) | 0.29 |
| Incontinence impact | –6.9 (2.3) | –22.3 (1.7) | <0.0001 | –7.2 (2.2) | –22.5 (1.7) | <0.0001 |
| Role limitations | –2.4 (2.1) | –17.3 (1.6) | <0.0001 | –4.2 (2.1) | –16.4 (1.6) | <0.0001 |
| Physical limitations | –0.4 (2.0) | –16.0 (1.6) | <0.0001 | –2.1 (2.0) | –15.4 (1.6) | <0.0001 |
| Social limitations | 1.4 (1.6) | –10.5 (1.3) | <0.0001 | –0.2 (1.6) | –9.9 (1.3) | <0.0001 |
| Personal relationships | –1.2 (2.0) | –7.7 (1.5) | 0.009 | –2.0 (2.0) | –7.1 (1.6) | 0.05 |
| Emotions | –1.6 (1.7) | –13.0 (1.3) | <0.0001 | –2.0 (1.7) | –13.2 (1.3) | <0.0001 |
| Sleep/energy | –0.1 (1.8) | –21.0 (1.4) | <0.0001 | –0.8 (1.8) | –20.8 (1.4) | <0.0001 |
| Severity measures | –1.1 (1.3) | –8.5 (1.0) | <0.0001 | –2.5 (1.3) | –7.7 (1.0) | 0.001 |

†Post hoc *t*-tests of least squares means between the two groups were performed.

change scores, except General Health Perception and Personal Relationships, were higher for those willing to continue treatment compared with those unwilling to continue treatment (Table 7).

### *Correlations*

Large, significant correlations demonstrating the high intercorrelation of these three concepts were found among all three BSW measures. The correlation between the benefit and satisfaction measures was 0.78 and 0.81 for Studies 1 and 2, respectively; the correlation between the benefit and willingness items was 0.74 for both studies; and the correlation between the satisfaction and willingness items was 0.73 and 0.78 for Studies 1 and 2, respectively (all $P < 0.001$).

The concurrent validity of the each of the BSW measures was evaluated using correlations with change from baseline to week 12 scores for micturition variables as well as OAB-q and KHQ quality of life domains for all available variables in the three studies (Table 8). Each of the BSW measures were moderately correlated with the micturition variables and all correlations were in the expected direction. Correlations ranged from –0.40 to –0.44, –0.33 to –0.44, and –0.29 to –0.40 for treatment benefit, satisfaction with treatment and willingness to continue treatment measures, respectively (Table 6). With the exception of the Social Interaction subscale, correlations among the BSW measures and the OAB-q were in the moderate to strong range and in the expected direction, ranging from 0.33 to 0.55, 0.33 to 0.53, and 0.28 to 0.46 for treatment benefit, satisfaction with treatment and willingness to continue treatment measures, respectively (Table 8). Correlations among the BSW measures and the KHQ were weak to moderate in strength and in the expected direction. With the exception of the General Health Perceptions domain, the correlations ranged from –0.14 to –0.45, –0.17 to –0.44, and –0.14 to –0.38 for treatment benefit, satisfaction with treatment and willingness to continue treatment measures, respectively (Table 8).

## Discussion

Symptom-based conditions require appropriate, validated PRO measures to assess treatment effects. Ideally, these measures should be short, easy to complete, easy to interpret, and clinically meaningful. Treatment benefit, satisfaction, and willingness to continue a treatment are all attributes that a successful treatment should possess. These measures used alone or in combination are valuable in informal clinical assessments and as end points in research studies. Rarely, however, have they been formally validated for these uses. In contrast, the BSW is a valid outcome measure that demonstrated consistency with both clinical outcomes and other PRO measures across three separate OAB clinical trials with varying patient populations.

Patient global assessments of treatment are similar to clinician global assessments of treatment benefit—except the point of view has changed. With the BSW, there is an inherent assumption that the patient weighs the risks and benefits of a treatment to provide a global response. When considering benefit, satisfaction, and willingness to continue, patients balance such issues as amount of symptom relief, life impact, side effects, drug costs, and convenience to provide their response. Although it may appear redundant to ask all three BSW questions, each question assesses a unique aspect of treatment outcome. For example, one may perceive a treatment benefit but not be satisfied with treatment due to overriding side effects or dosing inconvenience. Alternatively, one may perceive a benefit and be sat-

**Table 8** Spearman's correlations from all three studies

| Micturition measures | Patient perception of treatment benefit† | | | Patient satisfaction with treatment | | Patient willingness to continue with treatment | |
|---|---|---|---|---|---|---|---|
| | Study 1 | Study 2 | Study 3 | Study 1 | Study 2 | Study 1 | Study 2 |
| Micturition frequency | –0.41 | –0.40 | –0.32 | –0.35 | –0.41 | –0.34 | –0.36 |
| Mean number of nocturia episodes per night | –0.41 | –0.44 | na | –0.33 | –0.44 | –0.29 | –0.40 |
| Urgency Levels | –0.41 | –0.38 | na | –0.35 | –0.38 | –0.34 | –0.33 |
| Mean number of incontinence episodes per 24 h | na | na | –0.30 | na | na | na | na |
| OAB-q | | | | | | | |
| Symptom bother | –0.55 | –0.49 | na | –0.46 | –0.45 | –0.44 | –0.32 |
| Coping | 0.40 | 0.33 | na | 0.33 | 0.45 | 0.28 | 0.28 |
| Concern | 0.48 | 0.36 | na | 0.41 | 0.44 | 0.41 | 0.36 |
| Sleep | 0.59 | 0.54 | na | 0.50 | 0.49 | 0.46 | 0.42 |
| Social interaction | 0.24 | 0.09* | na | 0.24 | 0.15* | 0.22 | 0.05* |
| Total scale | 0.55 | 0.46 | na | 0.46 | 0.53 | 0.43 | 0.39 |
| KHQ | | | | | | | |
| General health perception | na | –0.12 | –0.09 | na | –0.11 | na | –0.04* |
| Incontinence impact | na | –0.32 | –0.35 | na | –0.29 | na | –0.26 |
| Role limitations | na | –0.23 | –0.31 | na | –0.24 | na | –0.21 |
| Physical limitations | na | –0.28 | –0.32 | na | –0.30 | na | –0.23 |
| Social limitations | na | –0.30 | –0.26 | na | –0.30 | na | –0.24 |
| Personal relationships | na | –0.19 | –0.14 | na | –0.17 | na | –0.16 |
| Emotions | na | –0.28 | –0.24 | na | –0.29 | na | –0.26 |
| Sleep/energy | na | –0.45 | –0.20 | na | –0.44 | na | –0.38 |
| Severity measures | na | –0.23 | –0.34 | na | –0.24 | na | –0.14 |

*Correlations were not statistically significant at the $P < 0.05$.
†Higher score of Benefit Satisfaction and Willingness to Continue reflect higher Benefit, Satisfaction, and Willingness to Continue. OAB-q scoring: Higher symptom bother scores indicate higher ratings of Symptom Bother whereas higher HRQL subscale scores indicate better HRQL. KHQ scoring: Higher values reflect greater HRQL impairment.
na, not assessed.

isfied because the balance of safety and efficacy is favorable, but not be willing to continue a treatment because the incremental benefit may not be enough to merit continuation or the cost may be prohibitive. Thus, responses to each question provide important information regarding perceived outcome and behavioral intention that is needed to assess treatment effectiveness, patient impact and future compliance with therapy. Furthermore, a negative response to one or more of the questions provides an opportunity to explore the reasons in greater detail.

The response options selected for each of the BSW measures demonstrated acceptable properties when evaluated using all levels or when collapsed into dichotomous categories. When Satisfaction, for example, was analyzed using all four responses, it was interesting to find that although the discrimination between "very satisfied" and "a little satisfied" was very clear, the discrimination between "a little satisfied" and "a little dissatisfied" was much less evident. Thus, the polar ends of the Satisfaction and Willingness scales were highly discriminative, whereas the middle options were more of a "gray area" for patients. The data were also analyzed as a dichotomous response by combining the similar responses (i.e., very satisfied/a little satisfied vs. very dissatisfied/a little dissatisfied) which yielded greater discrimination for both the Satisfaction and Willingness items.

As with any new measure, there are limitations to consider. For example, the implications of the unbalanced nature of the Benefit response scale have not been fully explored. Although the scaling is weighted toward positive responses with two responses for benefit and one for no benefit, there may be some rationale to ask how great the failure if no benefit is perceived. In addition, as with any interviewer-administered measure, there is always the possibility that patients will try to provide a socially desirable response. Nevertheless, given the corroborating evidence of consistent clinical and PRO outcomes with the BSW responses, there is little evidence of such a bias. Furthermore, all three studies include patients with a verified diagnosis of OAB; the two studies with all three BSW items were exclusive to patients with OAB and nocturia and the third trial required incontinence as a symptom for inclusion. Because of these criteria, the generalizability of the Satisfaction and Willingness to Continue items to all OAB patients may be lessened though the evidence from these studies suggests that they are robust. Further research will be required to determine if the BSW measures per-

form as well in other diseases or conditions, particularly in those with validated, multi-item satisfaction measures. In addition, it would be interesting to determine if the BSW results are affected by the order in which the questions are asked and interviewer versus self-administered format.

The BSW is a useful global measure to evaluate patient perception of treatment outcome. It does not, however, eliminate the need or value of multi-item disease-specific measures that provide more depth and insight into the multiple facets of complex constructs such as HRQL and treatment satisfaction. Assessing perceived benefit is not the same as assessing HRQL, nor is the use of the satisfaction with treatment item of the BSW the same as using a multi-item Satisfaction scale [16]. Global impression measures and multi-item measures each have their own place as PROs with strengths and weaknesses for both measures. In situations where the effects of a condition are varied and complex, multi-item measures will demonstrate greater reliability, validity, and responsiveness than single-item measures [17]. Single-item, global measures are short, easy to administer and to interpret and may be useful in conditions with multiple and varied symptoms (e.g., OAB), and in circumstances where the construct is an evaluation based on personal criteria that are not well understood or that vary from patient to patient (e.g., overall satisfaction). The choice between single- and multi-item measures requires a careful consideration of these factors.

The BSW is a promising tool for both the clinical and research settings to assist in assessing and managing treatment outcomes. It is a useful global impression of three key elements of treatment outcome: if there was a perceived benefit, if the treatment was satisfactory, and if the patient is willing to continue the treatment. All three issues, although related, capture different aspects of the patient's perception of a treatment. In addition to informing researchers in the development of new treatments, gaining an understanding of patient level perceptions can facilitate patient–physician communication and may increase adherence to prescribed treatment regimens in practice.

The authors would like to acknowledge Christine Thompson and Terra Slaton, MS, for their programming assistance and the subjects and investigators from the three studies who contributed data for this analysis.

Source of financial support: Pfizer Corporation.

## References

1 Revicki DA, Zodet MW, Joshua-Gotlib S, et al. Health-related quality of life improves with treatment-related GERD symptom resolution after adjusting for baseline severity. Health Qual Life Outcomes 2003;1:73. (Available at http://www.hqlo.com/content/1/1/67)
2 Colman SS, Brod MI, Krishnamurthy A, et al. Treatment satisfaction, functional status, and health-related quality of life of migraine patients treated with almotriptan or sumatriptan. Clin Ther 2001;23:127–45.
3 Coyne KS, Revicki DA, 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.
4 Leidy NK, Revicki DA, Geneste B. Recommendations for evaluating the validity of quality of life claims for labeling and promotion. Value Heath 1999;2:113–27.
5 Abrams P, Cardozo L, Fall M, et al. The standardization of terminology in lower urinary tract function: report from the standardization subcommittee of the international continence society. Urology 2003;61:37–49.
6 Kelleher CJ, Cardozo LD, Toozs-Hobson PM. Quality of life and urinary incontinence. Curr Opin Obstet Gynecol 1995;7:404–8.
7 Liberman JN, Hunt TL, Stewart WF, et al. Health-related quality-of-life among adults with symptoms of overactive bladder: results from a US community-based survey. Urology 2001;57:1044–50.
8 Coyne K, Matza L, Thompson C. The responsiveness of the Overactive Bladder Questionnaire (OAB-q). Qual Life Res 2005;14:849–55.
9 Studenski S, Hayes RP, Leibowitz RQ, et al. Clinical global impression of change in physical frailty: development of a measure based on clinical judgment. J Am Geriatr Soc 2004;52: 1560–6.
10 Huq ZU; RIS-GBR-32 Investigators. A trial of low doses of risperidone in the treatment of patients with first-episode schizophrenia, schizophreniform disorder, or schizoaffective disorder. J Clin Psychopharmacol 2004;24:220–4.
11 Monsch AU, Giannakopoulos P; GAL-SUI Study Group. Effects of galantamine on behavioural and psychological disturbances and caregiver burden in patients with Alzheimer's disease. Curr Med Res Opin 2004;20:931–8.
12 Foldemo A, Ek AC, Bogren L. Needs in outpatients with schizophrenia, assessed by the patients themselves and their parents and staff. Soc Psychiatry Psychiatr Epidemiol 2004;39:381–5.
13 Bonomi A, Cella D, Hahn E, et al. Multilingual translation of the Functional Assessment of Cancer Therapy (FACT) quality of life measurement system. Qual Life Res 1996;5:309–20.

14 Reese PR, Pleil AM, Okano GJ, Kelleher CJ. Multinational study of reliability and validity of the King's health questionnaire in patients with overactive bladder. Qual Life Res 2003;12:427–42.
15 Kelleher CJ, Cardozo LD, Khullar V, Salvatore S. A new questionnaire to assess the quality of life of urinary incontinent women. Br J Obstet Gynaecol 1997;104:1374–9.
16 Atkinson MJ, Sinha A, Hass SL, et al. Validation of a general measure of treatment satisfaction, the Treatment Satisfaction Questionnaire for Medication (TSQM), using a national panel study of chronic disease. Health Qual Life Outcomes 2004;2:12. (Available at http://www.hqlo.com/content/1/1/67)
17 Sloan J, Aaronson N, Cappelleri J, et al. Assessing the clinical significant significance of single items relative to summated scores. Mayo Clin Proc 2002; 77:479–787.

## Appendix A

*Benefit, Satisfaction, and Willingness to Continue (BSW) Questions*

The following questions are administered by the physician:
BENEFIT—please ask the patient the following question:
1. Have you had any benefit from your treatment? ☐ (0) No ☐ (1) Yes
If YES, please ask the patient the following question:
Have you had little benefit from your treatment or much benefit? ☐ (1) Little benefit ☐ (2) Much benefit

SATISFACTION—Please ask the patient the following question:
1. Taking all things into account, are you satisfied with your treatment?
☐ (1) Yes If YES, please ask the patient the following question:
Are you a little satisfied with your treatment or very satisfied with your treatment? ☐ (1) A little satisfied ☐ (2) Very satisfied
☐ (0) No If NO, please ask the patient the following question:
Are you a little dissatisfied with your treatment or very dissatisfied with your treatment? ☐ (1) A little dissatisfied ☐ (2) Very dissatisfied

WILLINGNESS TO CONTINUE
2. Would you be willing to continue treatment with this medication?
☐ (1) Yes If YES, please as the patient the following question:
Would you be a little bit willing to continue treatment with this medication or very willing to continue treatment with this medication? ☐ (1) A little bit willing ☐ (2) Very willing
☐ (0) No If NO, please ask the patient the following question:
Would you be a little bit unwilling to continue treatment with this medication or very unwilling to continue treatment with this medication? ☐ (1) A little unwilling ☐ (2) Very unwilling