Validation of the Patient Perception of Intensity of Urgency Scale in Patients with Lower Urinary Tract Symptoms Associated with Benign Prostatic Hyperplasia

Susan D. Mathias, MPH1–2*, Ross D. Crosby, PhD1,2,3, Jameel Nazir, MSc4, Monique Klaver, PhD5, Ted Drogendijk, MSc5, Zalmai Hakimi, PharmD5, Isaac A. Odeyemi, PhD4

1Health Outcomes Solutions, Winter Park, FL, USA; 2Neuropsychiatric Research Institute, Fargo, ND, USA; 3University of North Dakota School of Medicine and Health Sciences, Grand Forks, ND, USA; 4Astellas Pharma Europe Ltd., Chertsey, UK; 5Astellas Pharma Global Development, Leiden, The Netherlands

A B S T R A C T

Objective: To assess the reliability and validity of scores derived from the Patient Perception of Intensity of Urgency Scale (PPIUS) in patients with lower urinary tract symptoms (LUTS) associated with benign prostatic hyperplasia (BPH).

Methods: A post hoc analysis of the phase II Solifenacin and Tamsulosin in Males with Lower Urinary Tract Symptoms Associated with Benign Prostatic Hyperplasia trial (NCT00510406), a 12-week clinical trial in men with LUTS associated with BPH, assessed the measurement properties of six PPIUS-derived scores: mean score; maximum urgency score; total urgency and frequency score (TUFS; average sum of urgency scores over 3 days); and numbers of urgency episodes, urgency episodes of grade 3 or 4, and urgency incontinence episodes. Test-retest reliability, presence of floor/ceiling effects, responsiveness to change, known-group validity, and concurrent validity were assessed for each score. Results: A total of 901 patients had at least one valid PPIUS assessment after baseline. TUFS demonstrated good test-retest reliability (intraclass correlation coefficient >0.8), discriminated between groups defined based on International Prostate Symptom Score storage score severity (known-groups validity), had high concurrent validity, and had high responsiveness to change (Guyatt’s responsiveness statistic 0.88), with an absence of floor or ceiling effects. The psychometric properties of other PPIUS-derived scores were not as consistently robust and showed either low-to-moderate responsiveness, presence of a floor or ceiling effect, or low-to-moderate test-retest reliability.

Conclusions: This study shows that the PPIUS is reliable and valid in patients with LUTS associated with BPH. TUFS provided the best combination of psychometric properties of the six scores derived from the PPIUS and appeared to be an appropriate measure of urgency and frequency.

Keywords: benign prostatic hyperplasia, lower urinary tract symptoms, Patient Perception of Intensity of Urgency Scale, validation.

Introduction

Lower urinary tract symptoms (LUTS) include storage symptoms, such as increased daytime urinary frequency, nocturia, urinary urgency, and urinary incontinence; voiding symptoms, including weak stream, hesitancy, and terminal dribble; and postmicturition symptoms, which include incomplete bladder emptying and postmicturition dribble [1]. In a large population-based study (Epidemiology of LUTS) conducted in Sweden, the United Kingdom, and the United States, the prevalence of LUTS, defined as a symptom frequency of at least “A few times per week,” was more than 45% [1]. In men, LUTS are commonly associated with benign prostatic hyperplasia (BPH) [2,3].

Patients are typically in the best position to identify and describe their symptoms; therefore, the prevalence and impact of LUTS are often investigated using information collected directly from patients, through either surveys and questionnaires [4,5] or assessment tools including diaries or logs [6–8]. LUTS have been shown to negatively affect health-related quality of life (QOL) [9–11] because symptoms can cause significant interference with daily activities [12] and are associated with depression [11] and decreased enjoyment of sexual activity [13]. Reliable and valid tools for collecting information and assessing LUTS are essential. The Patient Perception of Intensity of Urgency Scale (PPIUS) is a single-item, patient-reported scale that can be used to assess the intensity of urgency associated with each micturition or incontinence episode (Table 1) [14]. The PPIUS uses information that patients provide regarding the frequency and urgency of micturition and incontinence episodes, and several scores can be derived from it. It was developed on the basis of guidance from the European Medicines Agency and then validated and used in clinical trials of therapies for overactive bladder (OAB), including those of solifenacin and mirabegron [8,14–22]. There are no reports, however, of its measurement properties in men with LUTS associated with BPH. The objective of this study was to assess the reliability and validity of different

* Address correspondence to: Susan D. Mathias, Health Outcomes Solutions, PO Box 2343, Winter Park, FL 32790, USA.
E-mail: smathias@healthoutcomessolutions.com.
1098-3015/36.00 – see front matter Copyright © 2014, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.jval.2014.09.002
scores derived from the PPIUS in a population of patients with LUTS associated with BPH.

Methods

Study Design

A post hoc analysis of PPIUS data from a phase II, double-blind, parallel-group, placebo-controlled, randomized study to assess the safety and efficacy of combinations of solifenacin and tamsulosin in men with LUTS associated with BPH (the phase II Solifenacin and Tamsulosin in Males with Lower Urinary Tract Symptoms Associated with Benign Prostatic Hyperplasia trial) was conducted. Full methodology, efficacy, and safety data from this trial have been reported previously [23]. Men aged at least 45 years diagnosed as having LUTS associated with BPH for at least 3 months, with a total International Prostate Symptom Score (IPSS) of at least 13 (a cutoff widely used to select patients for clinical trials of therapies for LUTS associated with BPH [23–26]), and a maximum urinary flow rate of 4 to 15 ml/s were eligible to enroll in the study. Patients with significant urinary, cardiovascular, cerebrovascular, renal, or hepatic disease were excluded.

Assessments

Assessments completed by patients enrolled in the study included the PPIUS; the IPSS [24], which, in addition to seven symptom items, contains a single-item assessment of QOL due to urinary symptoms (IPSS QOL); and the single-item Patient Perception of Bladder Condition tool. Each assessment was self-administered by the patient during clinic visits at screening, at the end of a 2-week run-in period (baseline), and at weeks 2, 4, 8, and 12 of treatment or at early termination.

Patients were asked to complete a 3-day micturition diary, in which they recorded all micturition and incontinence episodes, at home, before every visit. For every micturition and incontinence episode during the 3-day period, patients rated the degree of urgency according to the PPIUS, a five-point categorical scale, ranging from “No urgency” (a score of 0) to “Severe urgency” (a score of 3) and “Urgency incontinence” (4). Six scores derived from the PPIUS were evaluated (Table 1). The proportion of responses for each urgency rating was also recorded.

Patients completed the seven-item IPSS [27], developed to diagnose and assess the symptoms of BPH, at every visit. The IPSS was developed by the Measurement Committee of the American Urological Association and is recommended by the World Health Organization for use with patients with LUTS suggestive of BPH. It includes seven questions relating to the frequency of the following symptoms: incomplete bladder emptying, intermittency, weak stream, hesitancy, frequency, urgency, and nocturia. Each question is scored using a range from 0 to 5, where higher scores reflect increased symptom severity. The IPSS storage subscore and individual urgency item score were also calculated in this study. The IPSS also includes a single QOL question regarding how patients perceive their urinary condition, with responses ranging from “Delighted” (0) to “Terrible” (6).

Patients also completed the single-item Patient Perception of Bladder Condition, which asks patients to assess the amount of bother their urinary condition causes, using response options ranging from “Does not cause me any problems at all” (1) to “Causes me many severe problems” (6).

Descriptive Analysis

Patients with valid PPIUS assessments, that is, the PPIUS was completed correctly and PPIUS scores were calculated, at baseline were included in the analysis. Descriptive analysis was used to report demographic characteristics of the cohort including age, height, weight, and body mass index (BMI). Counts and percentages were calculated for race categories, history of smoking, and alcohol use.

Analysis of Measurement Properties

Measurement properties assessed included examination of response characteristics, test-retest reliability, responsiveness, known-groups validity, and concurrent validity. Internal consistency reliability could not be calculated because the PPIUS contains only a single item, and it was not possible to estimate the minimal clinically important difference because there were no anchors to which changes in PPIUS scores could be related. Measurement properties were evaluated for the full sample and for a subgroup of patients who reported substantial storage symptoms at baseline. Substantial storage symptoms were defined as having a micturition frequency of at least eight episodes daily and at least two episodes of urgency of grade 3 or 4 on average per day over the 3-day diary period.

To examine the PPIUS response characteristics, the variability in responses, including the number and percentage of patients with the minimum (0) and maximum (4) possible scores, was determined separately for the 3 days of diary ratings before baseline and the end of treatment for all six scores (mean and maximum urgency score, TUFS, and numbers of episodes of urgency [PPIUS grade ≥1], PPIUS grade 3 or 4, and urgency incontinence episodes). In addition, the proportion of responses for each urgency rating was determined separately by visit. Baseline and end-of-treatment visits were selected because these time points correspond to those for establishing efficacy.

Test-retest reliability refers to the extent to which a measure yields the same results in repeated applications in an unchanged population. Test-retest reliability was assessed by calculating intraclass correlation coefficients (ICCs) [28] between daily ratings on the PPIUS for 3 consecutive days before baseline. An ICC of at
least 0.80 was considered to indicate good reliability. In addition, a pairwise Wilcoxon signed rank test was used to determine whether there were any significant differences in ratings over the 3 days.

The responsiveness of a measure reflects its ability to detect clinically important changes, even if the changes are small. This was evaluated using three metrics. The standard effect size [29] was calculated as the difference in means between baseline and the end of treatment, divided by the SD of the change score at baseline. The magnitude of the effect was based on Cohen’s definitions of small (0.20), medium (0.50), and large (0.80). The standardized response mean [30] was calculated as the difference in means between baseline and the end of treatment, divided by the SD of the change score. Finally, the responsiveness statistic [31] was calculated as the difference in means between baseline and the end of treatment, divided by the standard deviation (SD) at baseline. The magnitude of the effect was based on Cohen’s definition of small (0.20), medium (0.50), and large (0.80).

Known-groups validity is the extent to which a measure is able to discriminate between groups known to be clinically different. Two groups based on symptom severity were created using the median IPSS storage score at baseline and the end of treatment: patients at or below the median were placed into one group, whereas those above the median were placed in the other group. The known-groups validity of the PPIUS at baseline and the end of treatment was then evaluated by comparing PPIUS scores between severity groups. Comparisons at each time point were made using analysis of covariance controlling for age. Adjusted means and SDs were calculated separately by group for each PPIUS score. Partial $\epsilon^2$ values, which express the proportion of variance of the dependent variable accounted for by group membership, were also calculated for each score at each time point.

A measure is considered to have adequate concurrent validity if it can be shown to be associated with conceptually related criterion measures. The concurrent validity of the PPIUS was examined in several ways. First, Spearman and Pearson correlation coefficients between baseline values of PPIUS and each of the following baseline scores were calculated: IPSS storage subscore, IPSS urgency item score, IPSS QOL score, and Patient Perception of Bladder Condition. Second, Spearman and Pearson correlation coefficients between the change in baseline-to-end of treatment PPIUS scores and the change in each of the above variables during that same time period were calculated.

### Results

#### Patient Population and Descriptive Analysis

A total of 1010 patients with LUTS associated with BPH and baseline PPIUS scores were included in the study. All participants were men, and all but five were white. Their mean age was 65.6 ± 8.0 years (range 45–92 years), and their mean BMI was 27.7 ± 4.0 kg/m². Although BMIs ranged from 17.6 to 54.3 kg/m², 23.5% of the participants met the criterion for obesity (BMI > 30 kg/m²). Only 13% of the participants were current smokers, and 63.3% were current users of alcohol. A total of 901 patients had at least one valid PPIUS assessment after baseline. Most of these patients (n = 862, 96%) had complete follow-up data, and therefore, end-of-treatment PPIUS scores represent those at week 12. The remaining patients terminated the study early, and their end-of-treatment values are from assessments before the 12th week. A total of 398 patients (40%) were included in the subgroup analysis, with end-point data available for 360.

#### Measurement Properties

A summary of PPIUS response characteristics is presented in Table 2. The characteristics of the urgency incontinence episode scores in this particular population are noteworthy: a high percentage of participants reported the minimum possible score of 0 (i.e., the “floor”) at baseline (79%), reflected in the median being 0. This would suggest that measurement of urgency incontinence may not be sensitive for detecting changes over time in the overall population of men with LUTS associated with BPH. Assessing the change in this parameter from baseline only in those patients who have urgency incontinence at baseline may address the issue of the high baseline floor effect.

The test-retest reliability results for PPIUS scores reveal that the highest ICC values were obtained for the mean PPIUS score, which averaged above 0.90, followed by TDFS and episodes of urgency of grade 3 or 4, which both averaged above 0.80 (Table 3). ICC values for the remaining scores averaged below 0.80, suggesting that they are somewhat less stable over time. All scores, except the mean PPIUS and urgency incontinence episode scores, demonstrated significant changes from day 1 to 2, suggesting that patients required some time to become accustomed to the diary recording procedures. Even though these changes were statistically significant, they were relatively small in magnitude (<0.08), indicating that the changes would likely not affect the validation of the PPIUS in this sample.

When examining the responsiveness statistics (Fig. 1), three scores show notably higher values than do other scores across all

### Table 2 – PPIUS response characteristics.

<table>
<thead>
<tr>
<th>Response characteristic</th>
<th>Time point</th>
<th>n</th>
<th>Mean ± SD</th>
<th>Median</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean PPIUS score</td>
<td>Baseline</td>
<td>1010</td>
<td>1.81 ± 0.71</td>
<td>1.90</td>
<td>0–4.00</td>
</tr>
<tr>
<td></td>
<td>End of treatment</td>
<td>901</td>
<td>1.60 ± 0.74</td>
<td>1.71</td>
<td>0–3.90</td>
</tr>
<tr>
<td>Maximum PPIUS score</td>
<td>Baseline</td>
<td>1010</td>
<td>2.59 ± 0.81</td>
<td>2.67</td>
<td>0–4.00</td>
</tr>
<tr>
<td></td>
<td>End of treatment</td>
<td>901</td>
<td>2.20 ± 0.91</td>
<td>2.00</td>
<td>0–4.00</td>
</tr>
<tr>
<td>TDFS</td>
<td>Baseline</td>
<td>1010</td>
<td>18.70 ± 7.92</td>
<td>17.67</td>
<td>0–92.33</td>
</tr>
<tr>
<td></td>
<td>End of treatment</td>
<td>901</td>
<td>14.51 ± 9.26</td>
<td>13.33</td>
<td>0–92.00</td>
</tr>
<tr>
<td>Urgency episodes</td>
<td>Baseline</td>
<td>1010</td>
<td>9.22 ± 3.30</td>
<td>9.00</td>
<td>0–22.33</td>
</tr>
<tr>
<td></td>
<td>End of treatment</td>
<td>901</td>
<td>7.78 ± 3.41</td>
<td>8.00</td>
<td>0–25.00</td>
</tr>
<tr>
<td>Urgency episodes of grade 3 or 4</td>
<td>Baseline</td>
<td>1010</td>
<td>2.51 ± 3.08</td>
<td>1.33</td>
<td>0–18.67</td>
</tr>
<tr>
<td></td>
<td>End of treatment</td>
<td>901</td>
<td>1.51 ± 2.73</td>
<td>0.00</td>
<td>0–25.00</td>
</tr>
<tr>
<td>Urgency incontinence episodes</td>
<td>Baseline</td>
<td>1010</td>
<td>0.33 ± 1.10</td>
<td>0</td>
<td>0–12.33</td>
</tr>
<tr>
<td></td>
<td>End of treatment</td>
<td>901</td>
<td>0.16 ± 0.92</td>
<td>0</td>
<td>0–17.00</td>
</tr>
</tbody>
</table>

PPIUS, Patient Perception of Intensity of Urgency Scale; TDFS, total urgency and frequency score.
the responsiveness statistics. These are the maximum PPIUS score, TUFS, and urgency episodes. The maximum PPIUS score had the highest standard effect size and standardized response mean, whereas TUFS produced the highest value for the responsiveness statistic.

Known-groups validity was assessed using data from baseline and the end of treatment, grouping patients by severity on the basis of the IPSS storage scores at the given time point. PPIUS scores were significantly higher for patients with higher severity (Table 4). At both baseline and the end of treatment, partial $\epsilon^2$ values were highest for TUFS (baseline = 0.125; end of treatment = 0.205) and urgency episode grades 1 to 4 (baseline = 0.111; end of treatment = 0.194), indicating that these two scores demonstrate the greatest ability to discriminate between groups of differing severity.

Several PPIUS scores demonstrated significant associations with conceptually related measures, indicating concurrent validity.

---

### Table 3 – Test-retest reliability of PPIUS scores at baseline.

<table>
<thead>
<tr>
<th>Response characteristic</th>
<th>Days compared</th>
<th>Mean 1</th>
<th>Mean 2</th>
<th>ICC (95% CI)</th>
<th>P$^\dagger$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean PPIUS score</td>
<td>1 vs. 2</td>
<td>1.82</td>
<td>1.81</td>
<td>0.904 (0.892–0.914)</td>
<td>0.199</td>
</tr>
<tr>
<td></td>
<td>1 vs. 3</td>
<td>1.82</td>
<td>1.80</td>
<td>0.885 (0.871–0.898)</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>2 vs. 3</td>
<td>1.81</td>
<td>1.80</td>
<td>0.915 (0.904–0.924)</td>
<td>0.137</td>
</tr>
<tr>
<td>Maximum PPIUS score</td>
<td>1 vs. 2</td>
<td>2.63</td>
<td>2.58</td>
<td>0.751 (0.722–0.776)</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>1 vs. 3</td>
<td>2.63</td>
<td>2.56</td>
<td>0.748 (0.719–0.774)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>2 vs. 3</td>
<td>2.58</td>
<td>2.56</td>
<td>0.786 (0.762–0.809)</td>
<td>0.296</td>
</tr>
<tr>
<td>TUFS</td>
<td>1 vs. 2</td>
<td>19.05</td>
<td>18.47</td>
<td>0.836 (0.817–0.854)</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>1 vs. 3</td>
<td>19.11</td>
<td>18.59</td>
<td>0.821 (0.800–0.840)</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>2 vs. 3</td>
<td>18.51</td>
<td>18.59</td>
<td>0.837 (0.817–0.854)</td>
<td>0.565</td>
</tr>
<tr>
<td>Urgency episodes</td>
<td>1 vs. 2</td>
<td>9.34</td>
<td>9.11</td>
<td>0.787 (0.762–0.809)</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>1 vs. 3</td>
<td>9.36</td>
<td>9.23</td>
<td>0.754 (0.726–0.780)</td>
<td>0.193</td>
</tr>
<tr>
<td></td>
<td>2 vs. 3</td>
<td>9.12</td>
<td>9.23</td>
<td>0.798 (0.775–0.820)</td>
<td>0.041</td>
</tr>
<tr>
<td>Urgency episodes of grade 3 or 4</td>
<td>1 vs. 2</td>
<td>2.60</td>
<td>2.49</td>
<td>0.832 (0.812–0.850)</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>1 vs. 3</td>
<td>2.60</td>
<td>2.46</td>
<td>0.826 (0.805–0.844)</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>2 vs. 3</td>
<td>2.49</td>
<td>2.46</td>
<td>0.836 (0.817–0.854)</td>
<td>0.162</td>
</tr>
<tr>
<td>Urgency incontinence episodes</td>
<td>1 vs. 2</td>
<td>0.31</td>
<td>0.34</td>
<td>0.749 (0.721–0.775)</td>
<td>0.462</td>
</tr>
<tr>
<td></td>
<td>1 vs. 3</td>
<td>0.31</td>
<td>0.35</td>
<td>0.801 (0.778–0.822)</td>
<td>0.154</td>
</tr>
<tr>
<td></td>
<td>2 vs. 3</td>
<td>0.34</td>
<td>0.35</td>
<td>0.798 (0.774–0.819)</td>
<td>0.821</td>
</tr>
</tbody>
</table>

CI, confidence interval; ICC, intraclass correlation coefficient; PPIUS, Patient Perception of Intensity of Urgency Scale; TUFS, total urgency and frequency score.

* n for comparisons were as follows: day 1 vs. day 2: 1008; day 1 vs. day 3: 1006; day 2 vs. day 3: 1005.

† Based on Wilcoxon signed rank test.

---

Fig. 1 – Responsiveness of the PPIUS-derived scores at week 12. SES was calculated as the difference in means between baseline and the end of treatment, divided by the SD at baseline. Magnitude of effect was based on Cohen’s definitions of small (0.20), medium (0.50), and large (0.80). SRM was calculated as the difference in means between baseline and the end of treatment, divided by the SD of the change score. The responsiveness statistic was calculated as the difference in means between baseline and the end of treatment, divided by the SD of the change score for stable patients. Values of 0.20 and 0.50 were used to indicate “small” and “medium” changes, respectively. PPIUS, Patient Perception of Intensity of Urgency Scale; SES, standard effect size; SRM, standard response mean; TUFS, total urgency and frequency score.
scores and the concurrent measures were also all significant (P < 0.001). The strongest relationship was demonstrated by TUFs, with Spearman correlations ranging from 0.33 for both the IPSS urgency item and the IPSS QOL item scores to 0.47 for the IPSS storage subscore (data not shown). Correlations between the baseline to end-of-treatment changes in PPIUS scores and the concurrent measures were also all significant (P < 0.05), again with TUFs demonstrating the strongest association (range 0.22-0.33).

A summary of the PPIUS validity measures is presented in Table 5. Although several of the PPIUS scores demonstrate adequate reliability and validity, TUFs stands out as consistently performing better than most or all the other PPIUS scores for each measurement property. The maximum PPIUS score and the urgency episode score demonstrate moderate-to-high reliability and responsiveness, followed by the mean PPIUS score and episodes of urgency of grade 3 or 4, which perform moderately well. The urgency incontinence episode score has consistently poor responsiveness and validity in the overall population, with limited ability to distinguish between patients of varying symptom severity or to identify significant changes within patients.

The analysis was also performed for the subgroup of patients with substantial storage symptoms at baseline (n=398). As expected, the baseline floor effect for urgency incontinence in this population was lower than that in the overall population (56% vs. 79%); similar results were observed in relation to the baseline floor effect for urgency grade 3 or 4 (0% vs. 31%). Results of the PPIUS validity analysis in this subgroup of patients with substantial storage as well as voiding symptoms at baseline were generally consistent with those of the full sample of patients with LUTS associated with BPH (data not shown).

**Discussion**

LUTS associated with BPH is common in the adult male population, especially in those of advancing age. Having access to valid, reliable, and easy-to-use tools to measure symptoms for patients with LUTS is critical for both researchers and clinicians. Currently, the IPSS, a seven-item measure of symptom severity, is most widely used to assess symptoms [27,32]. In contrast, the PPIUS is a single-item scale and has been validated as a measure of the urgency of micturition and incontinence episodes in patients with OAB [14]. If shown to be reliable and valid in men with LUTS, the PPIUS has potential advantages over the IPSS in terms of its ease of use and the ability to derive various scores from it. The current study provides evidence for the reliability and validity of the PPIUS in assessing the degree of urgency associated with micturition and urgency incontinence episodes in men with LUTS associated with BPH. The PPIUS demonstrates adequate-to-good reliability, responsiveness to change, the ability to differentiate groups differing in severity, and agreement with conceptually related measures in this population.

Of all the PPIUS-derived scores, TUFs appears to provide the best combination of psychometric properties. This measure, which takes both the level of urgency and the frequency into account, thus including two key storage symptoms in a single storage parameter, demonstrates good reliability and responsiveness, is able to differentiate among groups of differing severity better than do other PPIUS scores, and has a stronger association with other conceptually related measures at both baseline and when assessing change from baseline. It should be noted, however, that urgency and episodes of urgency of grade 3 or 4 also perform relatively well in the population of men with LUTS associated with BPH, and may therefore continue to be relevant for use in studies in this population. The high floor effect for urgency episodes of grade 3 or 4, that is, the proportion of men with no or mild urgency at baseline, may have had an influence on the responsiveness of this measure in this study as compared with studies of patients with OAB, in which assessment of urgency episodes of grade 3 or 4 is a key measure of disease burden. Because floor effects make it harder to detect changes over time, the responsiveness observed for urgency episodes of grade 3 or 4 supports its relevance. The floor effect was reduced in the subgroup of patients with substantial storage symptoms as well as voiding symptoms at baseline, consistent with the effect seen in patients with OAB; TUFs also provided the best combination of psychometric properties in this subgroup.

The statistically significant reduction in mean PPIUS scores between days 1 and 2 of the micturition diary before baseline may suggest that steps could be taken to try to improve consistency. The idea of teaching patients how to complete a micturition diary has been previously explored, but with mixed results, as discussed by Cartwright et al. [19]. Alternatively, it may be that patients will provide a more accurate assessment of symptoms if given at least 1 day to practice recordings before the data are used for analysis.

Although this study provides a comprehensive examination of the measurement properties of the PPIUS in patients with LUTS associated with BPH, its limitations should be noted. First, although test-retest reliability typically requires multiple assessments of untreated patients 7 to 14 days apart, participants in this study were being treated, and the data best suited to examining test-retest reliability were obtained on 3 consecutive days while patients were receiving placebo run-in medication.
rather than on separate occasions while patients were not receiving therapy. Although not ideal, these measures did allow for an initial examination of reliability. Another limitation is that no anchors were available to establish a minimally important clinical difference; specifically, the correlation between the IPSS and the PPIUS was not uniformly high enough (\(>0.30\)) for the IPSS to be an appropriate anchor. Once one is established, however, it will enhance the ability to characterize interindividual and intraindividual changes.

In summary, this study uses data from a large phase II clinical trial to provide evidence of the reliability and validity of the PPIUS in assessing urgency in men with LUTS associated with BPH. Given the prevalence of LUTS worldwide, its association with age, and variations in practice patterns in the diagnosis and treatment of BPH, it is crucial to be able to assess LUTS reliably and accurately. This is particularly important because evidence shows that physicians may be underutilizing currently available clinical assessments for LUTS and that the storage component of LUTS is underrecognized in men with BPH\(^{[33,34]}\). The PPIUS is a single-item scale that is reliable and valid, and it can be used with confidence in future research and clinical settings to assess urinary urgency in men with LUTS and the impact of treatment. TUFs, which takes both the level of urgency and frequency into account, thus including two key storage symptoms in a single storage parameter, appears to provide the best combination of psychometric properties in men with LUTS associated with BPH.

### Acknowledgments

This study was sponsored by Astellas Pharma Europe Ltd. Medical writing was provided by Andy Noble of Bioscript Medical and was funded by Astellas Pharma Europe Ltd.

Source of financial support: Funding for this study was provided in full by Astellas Pharma Europe Ltd.

### References


