Validity and reliability of the patient’s perception of intensity of urgency scale in overactive bladder

Rufus Cartwright, Sushma Srirkrishna, Linda Cardozo and Dudley Robinson
Department of Urogynaecology, King’s College Hospital, London, UK
Accepted for publication 7 May 2010

What’s known on the subject? and What does the study add?
Improvement in urinary urgency is an important goal for patients with overactive bladder, and should be measured along with change in other key symptoms in overactive bladder clinical trials. Existing scales that measure urinary urgency during completion of a bladder diary have incomplete evidence of validity.

Used as part of a 3-day bladder diary the Patient’s Perception of Intensity of Urgency Scale has good test-retest reliability and responsiveness. It correlates with other measures of condition severity, and distinguishes well between patient groups. The scale should therefore be useful both in clinical practice and in research.

INTRODUCTION

Urgency, ‘the complaint of a sudden compelling desire to pass urine’ [1] is the key, defining symptom of the overactive bladder syndrome (OAB). It is urgency itself that leads to frequency, nocturia and urgency incontinence [2-4]. The symptom of urgency has a greater impact on health-related quality of life than the other symptoms of OAB, including urgency urinary incontinence. A majority of patients with OAB have no leakage episodes at all [5], so-called OABdry. In study populations that include OABdry patients, traditional measurements of incontinence

OBJECTIVE

To assess the measurement characteristics of the Patient Perception of Intensity of Urgency Scale (PPIUS) in patients with overactive bladder (OAB).

PATIENTS AND METHODS

• Adult women with at least a 3-month history of OAB. The design was a 4-week, double-blind, randomized, placebo-controlled trial of transdermal oxybutynin, with a 2-week placebo run-in and 8-week, open-label extension.
• Symptom improvement was assessed using 3-day bladder diaries incorporating the PPIUS, and disease-specific health-related quality of life was assessed using the King’s Health Questionnaire (KHQ). Convergent validity was shown by correlation with the KHQ, and other bladder diary variables. Known groups validity was assessed by comparison of baseline mean urge ratings, and urgency episode frequency for continent and incontinent patients, and by comparison with the same measures from a historical control group of 40 asymptomatic female volunteers.
• Between- and within-groups responsiveness was assessed using standardized effect sizes (Cohen's d and effect size r). Reliability was assessed for the two arms of the trial at different time points and intervals, using intraclass correlation (ICC) and a t-test for the difference between mean scores.

RESULTS

• In total, 96 women were randomized. Urgency episode frequency showed moderate correlation with total KHQ score (r = 0.500, P < 0.001) and with daytime and night-time voiding frequency.
• There were significant differences in continent and incontinent subgroups for mean urge ratings (difference in means, −0.61/void, P < 0.001), and urgency episodes (difference in means, −2.67 episodes/day, P < 0.001), as well as between OAB patients and normal controls (mean urge rating: difference in means 1.22 per void, P < 0.001; urgency episodes: difference in means 2.93 episodes/day, P < 0.001).
• Between-groups analysis of effect size found that urgency episode frequency (d = 0.679, r = 0.321) was more responsive than mean urge rating (d = 0.480, r = 0.233). In both subgroups, urgency episode frequency (d = 0.421–0.454, r = 0.206–0.222) had better within-groups responsiveness than mean urge rating.
• Urgency episodes (ICC, 0.65–0.81) were measured more reliably than urgency urinary incontinence episodes (ICC, 0.50–0.65).

CONCLUSIONS

• Assessment of urgency episodes using the PPIUS shows good reliability, excellent known groups validity, high responsiveness and convergence with subjective measures of severity.
• PPIUS is freely available, and should be useful in both clinical practice and research studies when assessing women with urgency, with or without urgency incontinence.

KEYWORDS

overactive bladder, urinary urgency, urge urinary incontinence, oxybutynin, anticholinergic, patient goals
The SR-BD compares bladder sensation with voided volume, and existing reports have considered both the mean voided volume at each level of urgency, and the distribution of void ratings [12,13]. Coyne et al. [14] explored the responsiveness of modal rating, mean rating, and the sum of ratings with the PPIUS, and concluded that the sum of ratings was most responsive. This approach has been extended to other novel urgency scales [19].

By contrast, Cardozo et al. [15] defined levels 3 and 4 as urgency episodes (without or with urge incontinence respectively), and used the mean frequency of these urgency episodes as their primary outcome measure. This analysis is conceptually easily understood, and allows separate calculation of urgency episode frequency and urgency urinary incontinence episode frequency. The interpretation of voids according to this scheme is shown in Fig. 1.

The aim of the present analysis was to compare the validity and reliability of mean urge rating per void, urgency episode frequency defined as in Cardozo et al. [15], and urgency urinary incontinence episode frequency as outcome measures for women with OAB, with or without urgency urinary incontinence.

**MATERIALS AND METHODS**

The data were collected as part of a 4-week, double-blind, placebo-controlled randomized trial of transdermal oxybutynin [16]. In total, 98 female participants with at least a 3 month history of urinary urgency were recruited. Two thirds had one or more episode of urgency urinary incontinence on baseline 3-day diary (OABwet), and one-third were continent (OABdry). After a 2-week placebo run-in period, 96 women were randomized to 4 weeks of treatment with transdermal oxybutynin, or matching placebo patches. After 4 weeks of double-blind treatment, participants could opt to continue for a further 8 weeks of open-label treatment with transdermal oxybutynin. Symptom improvement was assessed using 3-day bladder diaries incorporating the PPIUS, and disease-specific health-related quality of life was assessed using the King's Health Questionnaire [20,21].

**CONVERGENT AND DIVERGENT VALIDITY**

Convergent validity of the mean urge ratings was shown by correlation with baseline scores.

---

**TABLE 1** Wording of the Indevus Urgency Severity Scale, Sensation-Related Bladder Diary and Patient Perception of Intensity of Urgency Scale

<table>
<thead>
<tr>
<th>Indevus Urgency Severity Scale</th>
<th>Sensation-Related Bladder Diary</th>
<th>Patient Perception of Intensity of Urgency Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: NONE – no urgency</td>
<td>Grade 1 – no desire to void</td>
<td>0 – No urgency: I felt no need to empty my bladder but did so for other reasons</td>
</tr>
<tr>
<td>1: MILD – awareness of urgency, but it is easily tolerated and you can continue with your usual activity or tasks</td>
<td>Grade 2 – desire to void but voiding can be delayed for at least 30 min</td>
<td>1 – Mild urgency: I could postpone voiding for as long as necessary without fear of wetting myself</td>
</tr>
<tr>
<td>2: MODERATE – enough urgency discomfort that it interferes with or shortens your usual activity or tasks</td>
<td>Grade 3 – desire to void but voiding can not be delayed for more than 15 min</td>
<td>2 – Moderate urgency: I could postpone voiding for a short while without fear of wetting myself</td>
</tr>
<tr>
<td>3: SEVERE – extreme urgency discomfort that abruptly stops all activity or tasks</td>
<td>Grade 4 – desire to void but voiding can not be delayed for more than 5 min</td>
<td>3 – Severe urgency: I could not postpone voiding but had to rush to the toilet in order not to wet myself</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 – Urge incontinence: I leaked before arriving at the toilet</td>
</tr>
</tbody>
</table>

episode frequency may fail to capture clinically meaningful improvements. Accurate assessment of urgency in OAB clinical trials has therefore been considered a priority. However, because of its subjective nature, it is difficult, in either a clinical or research setting, to distinguish pathological urgency from the normal physiological urge or desire to void. There remains debate as to whether urgency lies on a continuum with urge, as implied by most existing urgency scales [6,7], or whether urgency is a distinct ‘maximal’ sensation [3,8].

One approach to improving the objectivity of the measurement of urgency has been to incorporate an urgency scale into a bladder diary. Such scales include the Indevus Urgency Severity Scale (IUSS) [9,10], the Sensation-Related Bladder Diary (SR-BD) [11–13] and the Patient Perception of Intensity of Urgency Scale (PPIUS). The PPIUS has been recommended for use by the European Medicines Agency, and has subsequently been employed in trials of tolterodine [14] and solifenacin [15]. The reliability and normal ranges have been established in healthy female volunteers [17], although other aspects of validity have not been reported. The IUSS, SR-BD and PPIUS contain very similar items. Four grades of urgency are assigned a number and each is linked to the time available before voiding.
for the nine domains of the KHQ and the total KHQ score. Convergent validity for the count of urgency episodes/day (including urgency with and without urgency urinary incontinence) was shown by correlation both with the domains and total score of the KHQ, and by correlation with baseline daytime and night-time voiding frequency. Divergent validity for both measures was assessed by correlation with age, body mass index (kg/m²) and parity.

KNOWN GROUPS VALIDITY

Known groups validity was assessed by comparison of baseline mean urge ratings, and urgency episodes for patients who were OABdry at baseline, patients who were OABwet at baseline, and by comparison with the same scores from all patients enrolled in the present study, with a previous cohort of 40 asymptomatic female volunteers [17]. The t-test was used to assess differences between OABdry and OABwet, and between OAB and normal. One-way ANOVA was used to compare all three groups.

RESPONSIVENESS

To assess between-groups responsiveness, a t-test was used, in addition to calculation of standardized effect sizes (Cohen’s d and effect size r) for each measure at 4 weeks, based on the pooled standard deviations from the transdermal oxybutynin and placebo-treated subgroups. To determine within-groups responsiveness, bladder diary variables were compared before and after the switch from placebo to transdermal oxybutynin, using the same measures. For the subgroup randomized to transdermal oxybutynin, this switch occurred between baseline and 4 weeks. For the subgroup initially randomized to placebo, this switch occurred between the 4- and 12-week assessments.

TEST–RETEST RELIABILITY

Test–retest reliability was examined for daytime voids, night-time voids, 24-h voiding frequency, daily urgency episodes, daily urgency urinary incontinence episodes, and the mean urge rating per void. Reliability was also assessed for the two arms of the trial at different time points and intervals. For participants randomized to placebo, baseline assessments (after 2 weeks of placebo treatment) were compared with the 4-week assessments (after 6 weeks of placebo treatment). For participants randomized to transdermal oxybutynin, 4-week assessments (after 4 weeks of transdermal oxybutynin treatment) were compared with the 12-week assessments (after 12 weeks of transdermal oxybutynin treatment). Reliability was assessed using intraclass correlation (ICC), and t-testing for the difference between mean scores. ICC ≥ 0.6 was defined a priori as the minimum acceptable.

RESULTS

SCALE VARIABILITY

The modal rating of voids was 2, with a similar distribution at all time points (Fig. 2).

CONVERGENT AND DIVERGENT VALIDITY

Mean urge ratings showed significant or almost significant correlations with all domains of the KHQ, with the exception of the Personal Relationships domain. Pearson’s coefficient for each domain varied in the range 0.24–0.39, with the total score showing a moderate correlation of 0.50 (P < 0.001). Urgency episode frequency/day showed the same pattern with correlations 0.23–0.46, and correlation with the total score of 0.52 (P < 0.001). Urgency episode frequency/day was also significantly correlated with the other objective measures of OAB severity; daytime frequency (r = 0.30, P = 0.019), and night-time frequency (r = 0.36, P = 0.005). The full results of convergent validity analyses are given in Table 2. No significant correlations were found for either measure with age, parity or body mass index (all r ≤ 0.2 to +0.2; all P > 0.1) showing some divergent validity.

KNOWN GROUPS VALIDITY

The use of t-tests showed highly statistically significant differences in OABwet and OAB
TABLE 2 Pearson's correlations between mean urge rating and urgency episode frequency, and other subjective and objective measures of overactive bladder (OAB) severity

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean urge rating/void</th>
<th>Urgency episode frequency/day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daytime voids/day</td>
<td>r = -0.10†, p = 0.451</td>
<td>r = 0.30, p = 0.019</td>
</tr>
<tr>
<td>Night-time voids/day</td>
<td>r = 0.21†, p = 0.100</td>
<td>r = 0.36, p = 0.005</td>
</tr>
<tr>
<td>24-h frequency/day</td>
<td>r = -0.01†, p = 0.955</td>
<td>r = 0.38, p = 0.003</td>
</tr>
<tr>
<td>GH domain</td>
<td>r = 0.26, p = 0.046</td>
<td>r = 0.37, p = 0.003</td>
</tr>
<tr>
<td>Il domain</td>
<td>r = 0.30, p = 0.020</td>
<td>r = 0.32, p = 0.012</td>
</tr>
<tr>
<td>RL domain</td>
<td>r = 0.30, p = 0.018</td>
<td>r = 0.33, p = 0.010</td>
</tr>
<tr>
<td>PL domain</td>
<td>r = 0.28, p = 0.031</td>
<td>r = 0.33, p = 0.009</td>
</tr>
<tr>
<td>SL domain</td>
<td>r = 0.38, p = 0.002</td>
<td>r = 0.45, p &lt; 0.001</td>
</tr>
<tr>
<td>PR domain*</td>
<td>r = 0.17, p = 0.291</td>
<td>r = 0.10, p = 0.538</td>
</tr>
<tr>
<td>E domain</td>
<td>r = 0.39, p = 0.002</td>
<td>r = 0.39, p = 0.002</td>
</tr>
<tr>
<td>S domain</td>
<td>r = 0.24, p = 0.059</td>
<td>r = 0.23, p = 0.074</td>
</tr>
<tr>
<td>SM domain</td>
<td>r = 0.37, p = 0.003</td>
<td>r = 0.46, p &lt; 0.001</td>
</tr>
<tr>
<td>Total KHO</td>
<td>r = 0.500, p &lt; 0.001</td>
<td>r = 0.52, p &lt; 0.001</td>
</tr>
</tbody>
</table>

E, emotions; GH, general health perception; Il, incontinence impact; PL, physical limitations; PR, personal relationships; RL, role limitations; S, sleep/energy; SL, social limitations; SM, severity measures. *Personal relationships domain, n = 40; †Mean urge rating expected to be unrelated to measures of voiding frequency, as mean score is calculated per void.

RESPONSIVENESS

Between-groups analysis of effect size found that urgency episode frequency (d = 0.679, r = 0.321) was more responsive than mean urge rating (d = 0.480, r = 0.233). Daytime voiding frequency and 24-h frequency were also significant discriminators with moderate effect sizes. The full results are presented in Table 4.

Within-groups responsiveness was assessed in the same way; for the transdermal oxybutynin subgroup between baseline and 4 weeks and, for the subgroup initially randomized to placebo, between 4 and 12 weeks. In both subgroups, urgency episode frequency (placebo: d = 0.421, r = 0.206; transdermal oxybutynin: d = 0.454, r = 0.222) was more responsive than mean urge rating. No other measure was associated with moderate or better effect sizes in both tests of within-group responsiveness.

RELIABILITY

Test–retest reliability for the placebo and transdermal oxybutynin subgroups is described in Tables 5 and 6, respectively. The present study was not powered for reliability estimates and, with only 59.3% of patients completing all diaries, confidence intervals for...
ICC are wide. Estimates for daytime voids (0.58–0.83), night-time voids (0.66–0.68) and 24-h frequency (0.82–0.91) lie within the range of previous estimates of 3-day bladder diaries, despite these results being drawn from a therapeutic study, with differential responses in both arms being expected to compromise reliability. Urgency episodes (0.65–0.81) were measured more reliably than urgency urinary incontinence episodes (0.50–0.65), which probably reflects the low rate of urgency urinary incontinence episodes at baseline in this sample (0.8 per day). No significant difference in means was observed for any bladder diary variable in these analyses. Overall, the results indicate acceptable reliability of all measures, although reliability would clearly be improved by use of a longer diary.

**DISCUSSION**

Mean urge rating per void and urgency episode frequency both show convergent validity with the KHQ, and excellent known groups validity in discriminating normal and OAB patients, as well as OABdry and OABwet. Urgency episode frequency shows superior between- and within-groups responsiveness, and better reliability than either mean urge or urgency urinary incontinence episode frequency. On this basis, urgency episode frequency should be the preferred measure of urgency symptom severity using the PPIUS.

The present study represents a comprehensive validation of the PPIUS, including OABdry, OABwet and a historical group of normal volunteers. The 3-day PPIUS bladder diary shows similar reliability to a 7-day PPIUS diary [17], a 3-day SR-BD diary [11], a 7-day IUSS diary [9], or any one of a number of diaries that do not include urgency assessment [22]. In comparison to other scales, the PPIUS has shown better responsiveness and known groups validity, and greater convergence with the KHQ as a subjective measure of disease severity.

The generalizability of the analysis is threatened by its setting in an randomized controlled trial. Although compliance with diary keeping was not particularly high, which is typical of clinical practice, patients who participate in clinical trials may be quite unrepresentative in other ways. Equally, although the PPIUS bladder diary has appeal as a tool for population-based studies, the data may not be applicable in this setting. The validity of the PPIUS as a diagnostic tool is not assured either because the results were not compared in a population of women with stress urinary incontinence. Participants were not asked to measure voided volumes, and the additional use of the PPIUS may be associated with an unacceptable burden.

Future work should reassess the validity of the PPIUS in population-based samples of men and women, as well as a sample of men and women presenting for secondary care. Because bladder diaries are already widely used, with such information available, the scale could serve as a standard measure of urinary urgency to allow comparisons and meta-analysis of different studies.

In conclusion, the PPIUS is a freely available, extensively validated scale for the measurement of urinary urgency during completion of a bladder diary. It shows good reliability and responsiveness, and should therefore be useful in clinical practice and research studies when assessing patients with urgency, with or without urgency incontinence.

**ACKNOWLEDGEMENTS**


**CONFLICT OF INTEREST**

All authors have acted as consultants (LC) or paid speakers (RC, SS, DR) for UCB Pharma.

**REFERENCES**


---

**TABLE 5 Reliability of bladder diary parameters between baseline and 4 weeks for the subgroup randomized to placebo**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intraclass correlation</th>
<th>95% CI</th>
<th>Difference in means</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daytime voids/day</td>
<td>0.83</td>
<td>0.67–0.92</td>
<td>–0.05</td>
<td>0.87</td>
</tr>
<tr>
<td>Night-time voids/day</td>
<td>0.68</td>
<td>0.42–0.83</td>
<td>–0.02</td>
<td>0.88</td>
</tr>
<tr>
<td>24-h frequency/day</td>
<td>0.82</td>
<td>0.60–0.92</td>
<td>–0.07</td>
<td>0.76</td>
</tr>
<tr>
<td>Urgency episodes/day</td>
<td>0.65</td>
<td>0.37–0.82</td>
<td>0.21</td>
<td>0.48</td>
</tr>
<tr>
<td>Urgency incontinence episodes/day</td>
<td>0.50</td>
<td>0.15–0.73</td>
<td>0.23</td>
<td>0.05</td>
</tr>
<tr>
<td>Mean urge rating/void</td>
<td>0.72</td>
<td>0.37–0.82</td>
<td>0.01</td>
<td>0.73</td>
</tr>
</tbody>
</table>

**TABLE 6 Reliability of bladder diary parameters between 4 and 12 weeks for the subgroup randomized to transdermal oxybutynin**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intraclass correlation</th>
<th>95% CI</th>
<th>Difference in means</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daytime voids/day</td>
<td>0.58</td>
<td>0.21–0.80</td>
<td>0.05</td>
<td>0.91</td>
</tr>
<tr>
<td>Night-time voids/day</td>
<td>0.66</td>
<td>0.34–0.85</td>
<td>–0.10</td>
<td>0.60</td>
</tr>
<tr>
<td>24-hour frequency/day</td>
<td>0.91</td>
<td>0.81–0.96</td>
<td>–0.05</td>
<td>0.93</td>
</tr>
<tr>
<td>Urgency episodes/day</td>
<td>0.81</td>
<td>0.59–0.92</td>
<td>–0.50</td>
<td>0.31</td>
</tr>
<tr>
<td>Urgency incontinence episodes/day</td>
<td>0.65</td>
<td>0.31–0.84</td>
<td>–0.49</td>
<td>0.09</td>
</tr>
<tr>
<td>Mean urge rating/void</td>
<td>0.69</td>
<td>0.37–0.86</td>
<td>0.02</td>
<td>0.72</td>
</tr>
</tbody>
</table>
5 Irwin DE, Milsom I, Kopp Z, Abrams P, Cardozo L. Impact of overactive bladder symptoms on employment, social interactions and emotional well-being in six European countries. BJU Int 2006; 97: 96–100
11 Naoemova I, De Wachter S, Wyndaele JJ. Comparison of sensation-related bladder diaries during filling cystometry are different according to urodynamic diagnosis. Neurourol Urodyn 2005; 24: 397
17 Carter K, Somaroo C, Chaikin DC. Reliability of the 24-h sensation-related bladder diary in women with urinary incontinence. Int Urogynecol J Pelvic Floor Dysfunct 2008; 19: 28

Correspondence: Rufus Cartwright, Department of Urogynaecology, King’s College Hospital, Denmark Hill, London SE5 9RS, UK.
e-mail: rufus.cartwright@gmail.com

Abbreviations: ICC, intraclass correlation; IUSS, Indevus Urgency Severity Scale; KHQ, King’s Health Questionnaire; OAB, overactive bladder; PPIUS, Patient Perception of Intensity of Urgency Scale; SR-BD, Sensation-Related Bladder Diary.