Developing and Validating the International Consultation on Incontinence Questionnaire Bladder Diary

Elizabeth Bright a,*, Nikki Cotterill a, Marcus Drake b, Paul Abrams a

a Bristol Urological Institute, Southmead Hospital, Bristol, UK; b School of Clinical Sciences, University of Bristol, UK

Abstract

Background: Despite the common use of urinary diaries to assess lower urinary tract symptoms (LUTS), a standardised validated diary does not exist. Objective: To develop a validated urinary diary, using the psychometric validation protocol used in previous International Consultation on Incontinence Questionnaire (ICIQ) modules. Design, setting, and participants: We invited 400 consecutive patients attending the urology department for assessment of LUTS to complete a urinary diary (developed and validated for content in a previous study), and the ICIQ Male or Female LUTS questionnaire. Outcome measurements and statistical analysis: To establish construct validity, the urinary diary was compared with known theories from published literature; to establish criterion validity, the diary was compared with questionnaire responses and/or urodynamic observations. Optimal diary duration was tested by comparing the 4-d diary against shorter durations. Patients completed a second diary after 2–3 wk for test-retest analysis, and a subset receiving sacral nerve stimulation completed the diary before and after treatment for analysis of responsiveness. A variety of statistical tests were used for different stages of the study. Results and limitations: The urinary diaries and ICIQ LUTS questionnaires were completed by 264 patients. Construct validity was established for two of three tested hypotheses. Criterion testing showed good agreement between questionnaire and diary recordings of nocturia ($\kappa = 0.653; p < 0.001; 92.2\%$) and incontinence ($\kappa = 0.351; p < 0.001; 64.5\%$), whereas good agreement ($\kappa = 0.378; p < 0.001; 69.2\%$) was observed between urodynamically proven incontinence and diary reports ($n = 104$). Diary recordings of urgency showed weak agreement with questionnaire responses ($\kappa = -0.215; p < 0.001; 36\%$) and urodynamic observations ($\kappa = -0.105; p = 0.256; 43.7\%$). The 3-d diary explained at least 94% of the total variance of the 4-d diary. A second diary for test-retest analysis was returned by 59 patients, demonstrating fair to excellent agreement (Spearman correlations: 0.49–0.88). Pre- and post-treatment analysis, on pilot testing, showed that the diary is responsive to change. Conclusions: Using the ICIQ psychometric validation methodology, a bladder diary was developed for the assessment of LUTS and shown to be valid, reliable, and responsive to change. The 3-d diary has been accepted as the ICIQ bladder diary. Patient summary: In this study, patients and clinicians developed and tested a diary in which patients can record their urinary symptoms. The resulting 3-d diary is called the ICIQ bladder diary and is available for adult men and women with urinary symptoms.

© 2014 European Association of Urology. Published by Elsevier B.V. All rights reserved.

* Corresponding author. Bristol Urological Institute, Southmead Hospital, Bristol, BS10 5NB, UK. Tel. +44 117 323 5690; Fax: +44 117 323 8830. E-mail address: ebright9@gmail.com (E. Bright).

http://dx.doi.org/10.1016/j.eururo.2014.02.057
0302-2838/© 2014 European Association of Urology. Published by Elsevier B.V. All rights reserved.
1. Introduction

Urinary diaries are commonly used, recommended tools in the investigation and management of patients with lower urinary tract symptoms (LUTS) [1,2]. However, although other patient-completed symptom assessments, such as questionnaires, have been validated for use in clinical and research settings, a standardised, validated urinary diary appears not to exist [3]. Numerous diaries displaying different content, format, and duration are currently available. The International Continence Society (ICS) has defined the following three types of urinary diary on the basis of the recorded parameters [4]:

- Micturition charts record only the times of micturition, day and night, for at least 24 h.
- Frequency-volume charts record the volumes voided and the time of each micturition, day and night, for at least 24 h.
- Bladder diaries record the times of micturition and voided volumes, as well as other information, such as incontinence episodes, pad usage, fluid intake, degree of urgency, and the degree of incontinence.

The objective was to develop and validate a urinary diary applicable to male and female adults, using the psychometric validation protocol required by regulatory authorities [5], applied to the most recent International Consultation on Incontinence Questionnaire (ICIQ) modules (see www.iciq.net) [6]. This study has been conducted in two phases. The first phase has been completed and published [7] and was concerned with gathering raw data from a variety of sources (topic exploration; phase 1a), with subsequent development and testing of a succession of diaries that underwent content validation (phase 1b). The resulting diary, of portrait format with a printed 24-h clock and 4-d duration, records urinary frequency, voided volumes, fluid intake (amount, time, and type), time of incontinence episode, and bladder sensation. The current paper describes further validation of this diary via tests of construct validity, criterion validity, reliability, responsiveness, and the evaluation of optimum diary duration (phase 2).

2. Materials and methods

Four hundred consecutive patients attending either the urodynamic (n = 200) or uroflowmetry clinic (n = 200) were requested to complete the diary and a sex-specific ICIQ LUTS questionnaire (ICIQ-MLUTS for men, and ICIQ-FLUTS for women) prior to attendance. Diaries and questionnaires were mailed to patients with their appointment letter and a patient information sheet detailing the study. No exclusions to recruitment were applied. There are no specific numbers required to undertake psychometric validation, but groups large enough to be representative are essential [8,9]. As there was no difference to be detected and no known effect size, a power calculation was impossible. As such, formal sample-size calculations could not be performed; studies of questionnaire design and validation often simply recruit as many participants as possible to be as representative of the population as feasible, with recruitment often based on previous literature. A large cohort of patients, therefore, was recruited to ensure a reasonable number of diary returns in both the first and second diaries.

2.1. Construct validity

Construct validity measures relationships between the diary and underlying theories. If the expected relationships are observed, this provides further evidence of diary validity. Published literature was used to generate the following hypotheses regarding LUTS: (1) Incontinence is more common in women than men, (2) the prevalence of nocturia increases with age, and (3) incontinence is more prevalent in older individuals than young adults.

Diary responses were analysed and compared to the hypotheses. The Mann-Whitney test and binary logistic regression were calculated to investigate incontinence and patient age. The Mann-Whitney test was used to compare nocturia and patient age. Chi-square analysis was used to assess sex and incontinence differences.

2.2. Criterion validity

Criterion validity measures the relationship between the new tool being developed and an accepted gold standard measure of the concept in question or another validated tool, such as the ICIQ LUTS questionnaires.

Hypotheses were generated regarding expected associations between diary data and ICIQ LUTS questions. Three parameters—incontinence, urgency, and nocturia—were suitable for evaluation, as these could be measured as either present or absent in both the diary and questionnaire. Further assessment was undertaken in the subgroup of patients that underwent urodynamic studies, allowing the comparison of urodynamic study (UDS) observations with diary-documented urgency and incontinence. Agreement between diary and questionnaire responses or UDS results were measured by kappa statistic and percentage agreement, with $\kappa > 0.4$ and agreement $>50\%$ deemed acceptable [10–12].

2.3. Evaluation of optimal diary duration

Mean measurements for all diary parameters were calculated for diary day 1, diary days 1 and 2, diary days 1–3, and for the entire 4-d duration. Using the 4-d average as the benchmark, the other durations were compared via Spearman correlations. To detect the presence of urgency and incontinence, the percentage of patients who documented the first episode of each symptom on diary day 1, 2, 3, or 4 were calculated.

2.4. Reliability (test-retest analysis)

Reliability measures the ability of the diary to produce consistent results. Test-retest analysis was measured by the administration of the diary on two occasions between which it is not expected that the condition will change (hence the diary should not detect a change), but which are sufficiently separated so that respondents cannot recall their initial responses.

All patients who returned a completed diary were requested to complete a second diary 2–3 wk later. Patients were handed a pack during clinic attendance containing a patient information sheet, a second diary, a study-specific questionnaire to ensure there had been no change to their medical treatment during the interim period, and a prepaid addressed envelope for return. To establish test-retest reliability, the average of each recorded diary parameter for each patient was calculated in both the first and second urinary diary and compared within subjects. Correlation coefficients and percentage agreement were calculated for continuous variables and nominal data, respectively.

2.5. Responsiveness

Responsiveness measures the ability of the diary to detect a real change over an expected period of time, such as before and after treatment.
A cohort of patients who received sacral nerve stimulation after completion of the diary provided a pilot study to evaluate responsiveness. Patients were mailed a pack containing a patient information sheet, a diary, and a prepaid addressed envelope for return, and were requested to complete the diary 2 wk after insertion of the test stimulator. Averages for each diary parameter for each patient in pre- and post-treatment diaries were calculated and analysed using the related t test for parametric continuous data, Wilcoxon test for nonparametric continuous data, and McNemar test for nominal data.

Ethical approval to undertake this study was granted by Southmead Research Ethics Committee. Statistical advice and analysis were provided by independent statisticians from the Peninsular Medical School, Exeter, UK, and the University of Plymouth, UK.

3. Results

Of the 400 patients requested to complete the diary and ICIQ LUTS questionnaire, 264 (66%) patients (123 women, 141 men; age range: 16–89 yr; mean age: 60.5 ± 16.7 yr) returned a completed diary and the ICIQ-MLUTS or ICIQ-FLUTS questionnaire. Diaries were completed prior to urodynamic and uroflowmetry investigation in 108 and 156 patients, respectively. Of the 264 patients, 214 completed all 4 d of the diary; the remainder completed 1 d (n = 31), 2 d (n = 15), and 1 d only (n = 4).

3.1. Construct validity

Significantly more women than men reported urinary incontinence (p < 0.001), and patients reporting nocturia were significantly older than their non-nocturic counterparts (p = 0.003). The expected increase in the prevalence of incontinence with age, however, was not observed (p = 0.045), with an odds ratio of 1.0036 (95% confidence interval, 0.9867–1.0208), indicating that a 1-yr increase in age minimally affects the probability of being incontinent. Evidence of construct validity was thus established in two of three tested hypotheses.

3.2. Criterion validity

Table 1 details the statistical associations between diary and questionnaire responses or UDS. Nocturia revealed strong associations, and incontinence demonstrated a reasonable percentage agreement but weak correlation; a poor percentage agreement and a weak correlation were observed for urgency.

3.3. Evaluation of optimum diary duration

The entire 4-d diary was completed by 81% of patients. Correlation values for diary parameters of different durations are shown in Table 2. The first 3 d of the diary, the first 2 d, and the first day explain at least 94%, 88%, and 71% of the total variance of the 4-d diary, respectively.

For patients reporting urgency, the percentage reporting their first episode increased with longer diary durations: 86.7%, 94%, and 98.2% for diary durations of 1, 2, and 3 d, respectively. This increase in percentage with longer diary durations was also observed for those patients reporting incontinence: 1-d diary, 69.4%; 2-day diary, 87.1%; and 3-d diary, 96.5%.

3.4. Reliability (test-retest analysis)

The completed second diary was returned by 59 patients (15 women, 44 men; age range: 27–89 yr; mean age: 69.3 ± 12.8 yr) 2–3 wk after completion of the first. No patients had received additional treatment during this period. The parameters recording urinary frequency, voided volumes, fluid intake, and the number of incontinence episodes all displayed good or excellent agreement on test-retest analysis. Levels of agreement varied from fair to excellent agreement for different bladder sensation scores (Table 3), and 82.8% and 78% agreement was observed between diaries for the presence or absence of urgency and incontinence, respectively.

Attrition analysis was performed by comparing several diary parameters on the initial diaries of patients who completed (n = 59) and patients who failed to complete (n = 205) a second diary for test-retest analysis. No
A statistically significant reduction in 24-h frequency, daytime frequency, nocturia, and incontinence episodes was observed after treatment. Although no significant reduction in the total number of patients reporting the presence of urgency was demonstrated, post-treatment bladder sensation scores did demonstrate less urgency overall. A significant increase in the number of voids recorded with a bladder sensation score of 1 (normal desire to void) with a corresponding reduction in the number of voids with a score of 3 (urgency) or score of 4 (urgency incontinence) was observed (Table 4).

### 4. Discussion

The aim of this study was to develop a validated urinary diary to be used in the assessment of male and female adult patients with LUTS. Although urinary diaries are commonly

<table>
<thead>
<tr>
<th>Diary parameter, no.</th>
<th>Pretreatment diary, median (IQR)</th>
<th>Post-treatment diary, median (IQR)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder sensation score 0</td>
<td>0 (0–0.25)</td>
<td>0 (0–0.5)</td>
<td>0.506</td>
</tr>
<tr>
<td>Bladder sensation score 1</td>
<td>0.25 (0–2)</td>
<td>3 (1–4.825)</td>
<td>0.002</td>
</tr>
<tr>
<td>Bladder sensation score 2</td>
<td>0.25 (0–1.25)</td>
<td>0.25 (0–0.75)</td>
<td>0.799</td>
</tr>
<tr>
<td>Bladder sensation score 3</td>
<td>2.75 (2.25–4.875)</td>
<td>1 (0.375–2.5)</td>
<td>0.01</td>
</tr>
<tr>
<td>Bladder sensation score 4</td>
<td>2.75 (1.25–4.25)</td>
<td>0 (0–1.375)</td>
<td>0.005</td>
</tr>
<tr>
<td>Leaks per 24 h</td>
<td>14 (5–17.5)</td>
<td>0 (0–5.5)</td>
<td>0.005</td>
</tr>
<tr>
<td>Leaks per 24 h</td>
<td>3.5 (1.25–4.375)</td>
<td>0 (0–1.375)</td>
<td>0.005</td>
</tr>
</tbody>
</table>

IQR = interquartile range; SD = standard deviation.

* p < 0.005.
used in both clinical practice and research, there is limited evidence in the published literature regarding their validation [3]. Furthermore, the use of psychometrically validated patient tools in research trials is now a requirement of regulatory authorities such as the US Food and Drug Administration [5].

The aim of the ICIQ is to develop validated tools for the assessment of pelvic dysfunction [6]. Consequently, the ICIQ advisory board recommended the development of a validated urinary diary to assess LUTS. From the outset, it was appreciated that it would be difficult to produce a diary that would satisfy every potential user or situation. A one size fits all diary was unlikely to be developed. However, the aim was to at least try to create a generic diary. By incorporating the views of both patients and clinicians at conception, the resulting diary at the end of phase 1 exhibited a format that was interpretable by patients and acceptable to clinicians, resulting in a consensus opinion on diary content. The 4-d diary was unlikely to be developed. However, the aim was to at least try to create a generic diary. By incorporating the views of both patients and clinicians at conception, the resulting diary at the end of phase 1 exhibited a format that was interpretable by patients and acceptable to clinicians, comprising a consensus opinion on diary content. The 4-d duration was retained to allow further analysis of optimal diary duration in a larger dataset in phase 2 [7].

In this phase of the study, evidence for diary validity, reliability, and responsiveness has been established. In the development of patient-reported outcome measures, such as self-completion symptom diaries and questionnaires, validation involves an ongoing process of gathering evidence for the establishment of validation [10,12]. Although different substudies are performed to explore evidence of validity, it is known that not all analyses may provide positive results. However, such negative findings do not negate other positive evidence observed for validity, and hence the ability to describe that tool as being valid [11,13]. Basic validation of patient-reported outcome measures requires evidence of content validity and reliability. Additional tests such as criterion validity, construct validity, and responsiveness are supplementary and, as such, are included to offer any further evidence of validation. Evidence for the diary’s construct validity was observed in two of three tested hypotheses. A methodological flaw is the likely reason for failing to observe that incontinence, as detected by the diary, increases with age. Incontinence rates in the cohort of patients attending the clinic were compared to epidemiologic, published community populations, encompassing patients with and without LUTS. As such, the study population represented a higher proportion of patients of all ages with incontinence, and it is appreciated that this does not provide a direct comparison. However, the percentage of patients reporting incontinence (40.9%) and urgency (70.8%) in the study dataset is considerably higher than that documented in the epidemiologic literature, as would be expected on comparison with a cohort of patients attending the urology clinic with LUTS [14,15]. Thus, further evidence for the construct validity of the diary has been observed.

Variable agreement was observed on criterion testing. Disagreement between diary and questionnaire documentation, for both incontinence and urgency, may be the result of poor patient recall during diary completion, the over- or underreporting of symptoms within the questionnaire, or the infrequent nature of some symptoms, such that it is not detected during a 4-d period. In addition, patients may adopt habits to prevent the occurrence of incontinence [16] or urgency day to day [12,17], such as voiding more frequently to prevent their bladder from reaching higher capacities. Although UDSs provide an objective assessment of patient symptoms, comparison with diary recordings of incontinence could be inaccurate because incontinence may not be reproduced during UDSs [18]. Similarly, regarding urgency, a lack of diary and UDS correlation may be because urgency only occurred at the extremes of bladder filling not assessed during UDSs or that the catheter used in a UDS caused urgency not usually experienced by that individual. Consequently, UDSs do not replicate real life in all patients [19,20]. Although some elements of criterion testing have shown less than anticipated correlations, this has been documented in previous studies of validity and, importantly, does not negate the positive evidence of criterion testing observed in this study [11,13].

Analysis of diary duration showed the 3-d version to be almost as reliable as the 4-d diary. Published literature has demonstrated a minimum of 3 d is required to achieve reliability for a variety of diary parameters [20–23]. A 3-d diary reduces patient burden, thus potentially enhancing patient compliance [24,25], and was the preferred duration of the clinicians surveyed during phase 1. Thus, the 3-d period appears optimal and is recommended for this diary.

Of the 264 patients, 59 returned a completed second diary for test-retest analysis, representing a considerable attrition rate. Analysis of symptom severity between the two cohorts of patients who completed and patients who failed to complete a second diary showed no statistically significant difference in several diary parameters. Consequently, the severity of patient symptoms was not felt to contribute to the smaller number of responders in this stage and, hence, was unlikely to bias the results of reliability testing. It is noted that the first diary was completed as part of a clinical appointment, whereas the second diary was for research purposes only, which may have affected the motivation for completion. In addition, the mean age of those who failed to complete a second diary was significantly younger and may represent a working population that did not have time to undertake the second diary, although this is only a postulation, as employment status was not recorded in the dataset. Although the sample size in this part of the study was much smaller than the original 264-patient dataset, it is similar to test-retest published literature of other psychometrically validated symptom tools [11,13,26,27]. Consequently, the overall reliability of the diary is not considered to be affected by the sample size or the skewed sex distribution within this sample. Good or excellent agreement was demonstrated on test-retest analysis for all parameters except bladder sensation scores of 2 (urgency subsided before voiding) and 4 (urgency incontinence). These scores were documented in the diaries less often and their infrequent use may have resulted in reduced reliability. However, in a population with a higher prevalence of overactive bladder, the full range of scores would be valuable. Therefore, although these weaker scores could be excluded, reducing the bladder sensation score to only three options (0, 1, and 3), this would weaken the score overall and it would be unable to present
the severity of urgency as a continuum [28]. Furthermore, it should be pointed out that the duration of the bladder sensation score is no longer than previously published urgency scales [12,29–33]. Although perfect agreement on test-retest correlation is highly unlikely due to natural fluctuations in bladder function, patient perceptions, and social situations during the different periods of diary completion, it is important to test consistency to ensure that the information gathered is as accurate as possible.

Although the majority of parameters performed well in validity and reliability analysis, urgency performed less well on criterion validity testing and two of the bladder sensation scoring components showed only fair agreement on test-retest analysis. The bladder sensation score, however, has shown good results in an initial pilot study of sensitivity to change analysis and, as such, the scoring system has shown evidence of validity in some stages of the study. The bladder sensation score, therefore, remains in the final diary, although additional testing in larger cohorts of different patient or treatment groups is planned to obtain further evidence and will be the basis of future research. Overall, however, in this small sample, the diary was responsive to expected changes following treatment. The final diary has a duration of 3-d, retains a portrait format with a printed 24-h clock, and records urinary frequency, voided volumes, fluid intake (amount, time, and type), time of incontinence episode, and bladder sensation. Following ICIQ advisory board review of this diary, an additional column to document incontinence pad use has been added. This was originally excluded, as it was not selected by patients and clinicians during content validation. However, it is appreciated that pad use is frequently recorded in diaries used in both clinical and research practice and its inclusion would enhance diary versatility. In the development of patient-reported outcome measures, existing tools are often borrowed and adapted for future use in different scenarios [12]. It is anticipated, therefore, that in future research, researchers may wish to add more complex items such as bladder pain scores or voiding symptom scores in the form of interchangeable bolt-on columns for use in different situations, thus providing a urinary diary for particular patient groups. However, if significant additions are made, then further study would be required to validate such additions prior to use. Finally, this diary has been validated using a British English-speaking population and as with all such tools, will require cultural adaptation and linguistic validation for other languages using the formal ICIQ protocol [6]. To prevent the contradiction of ICS terminology, the term urinary diary has been used throughout the project, as it was not known from the outset what would result. It can now be observed that the tool fits the ICS criteria for a bladder diary and, therefore, will be referred to as the ICIQ bladder diary (Supplemental Fig. 1).

5. Conclusions

This study describes the final development of the first validated bladder diary for the assessment of LUTS in both male and female adults, using a psychometric validation methodology. The resulting bladder diary is recommended for use over a 3-d period and has been accepted as the ICIQ bladder diary. The inclusion of the diary in current national research studies will provide ongoing evidence of validation, as well as the external validity of the diary.

Author contributions: Elizabeth Bright had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Bright, Cotterill, Drake, Abrams. Acquisition of data: Bright. Analysis and interpretation of data: Bright, Cotterill. Drafting of the manuscript: Bright, Cotterill, Drake, Abrams. Critical revision of the manuscript for important intellectual content: Bright, Cotterill, Drake, Abrams. Statistical analysis: None. Obtaining funding: None. Administrative, technical, or material support: Bright, Cotterill. Supervision: Cotterill, Drake, Abrams. Other (specify): None.

Financial disclosures: Elizabeth Bright certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: None

Funding/Support and role of the sponsor: None.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.eururo.2014.02.057.

References

Leidy NK, Vernon M. Perspectives on patient-reported outcomes: content validity and qualitative research in a changing clinical trial environment. Pharmacoeconomics 2008;26:363–70.


De Wachter S, Wyndaele JJ. How sudden is a compelling desire to void? An observational cystometric study on the suddenness of this sensation. BJU Int 2008;101:1000–3.


