**The development of the ICIQ-UAB: a patient reported outcome measure for underactive bladder**

**Abstract**

**Aims**

To present the development of the ICIQ -UAB as the first patient reported outcome (PRO) measure for the assessment of the symptoms and impact on health-related quality of life of underactive bladder (UAB), developed in-line with the Food and Drug Administration (FDA) Guidance for Industry.

**Methods**

Draft items were developed following 44 semi-structured concept elicitation interviews in the UK, and refined using 36 cognitive interviews. A pilot study was designed to assess the draft ICIQ-UAB's initial psychometric properties with 54 patients, recruited from European hospitals. Further concept elicitation interviews were also carried out with 11 patients in the US and 10 patients in Japan. All participants had a prior urodynamic diagnosis of detrusor underactivity.

**Results**

The cognitive interviews confirmed the initial items to be understood and interpreted as intended. Pilot testing showed that both internal consistency (Cronbach's $\alpha \geq 0.85$) and test-retest reliability (stable patients; intraclass correlation coefficient ≥0.88) was high. The interviews in the US and Japan elicited symptoms and impacts that support previous findings in the UK and provided further insight into the experiences of patients in those countries. The developmental ICIQ-UAB was refined using the evidence from all sub-studies.

**Conclusions**

The validity and reliability of the ICIQ-UAB was supported in a pilot study setting, and the wider cultural applicability by the additional interviews in the US and Japan. The developmental ICIQ-UAB is ready for further validation in future clinical trials and is envisaged as an important tool for the monitoring of future UAB treatment strategies.

**Key words:** Detrusor underactivity, patient reported outcomes, qualitative, quality of life, underactive bladder

## INTRODUCTION

Underactive bladder (UAB) is a health issue which is receiving increasing attention in the urological literature. Analogous to the relationship between overactive bladder (OAB) and urodynamic detrusor overactivity (DO), UAB is considered a symptom syndrome suggestive of the urodynamic observation of detrusor underactivity (DU) [1,2]. The most recent symptomatic definition approved by the International Continence Society steering committee in 2016 states: "Underactive bladder is characterised by a slow urinary stream, hesitancy and straining to void, with or without a feeling of incomplete bladder emptying and dribbling, often with storage symptoms" [3]. The epidemiology of DU is poorly understood [4], as it can only be diagnosed reliably by urodynamic testing, and there is considerable overlap of presenting lower urinary tract symptoms (LUTS) with the often co-existing urological conditions such as bladder outlet obstruction and DO [5,6]. However, DU is observed in up to 40% of men and 13% of women in those referred for urodynamic assessment [5], increasing to 48% in men over 70 years of age [7]. LUTS associated with UAB/DU can have a broad impact on health related quality of life for patients [8], in particular the impact of nocturnal voids, urinary tract infections, and the inconvenience of self-catheterisation [9].

Patient reported outcome measures (PROs) are in widespread use for the initial assessment, follow-up and monitoring of treatment strategies for lower urinary tract symptoms (LUTS). However, no PRO instruments currently exist, to assess UAB symptoms and impacts, that would meet the standards of validation described by the FDA's guidance for Industry [10]. The International Consultation on Incontinence Questionnaire (ICIQ) modules offer a range of psychometrically robust instruments for the self-assessment of lower pelvic dysfunction including LUTS [11]. The aim of this new module (the ICIQ-UAB) is to capture the patient reported symptoms of UAB, and their associated bother and impact, for use as an outcome measure in future clinical trials and clinical practice. This paper presents the results of several sub-studies for the development and initial validation of the ICIQ-UAB.

## MATERIAL AND METHODS

A flow diagram illustrating the developmental progression of the ICIQ-UAB to date is given in Figure 1. Country-specific ethics approval was granted for all sub-studies.

### *Urodynamic inclusion criteria*

*A priori* urodynamic inclusion criteria were essential to select adult, symptomatic patients with the urodynamically confirmed diagnosis of DU. Patients were also included who had common co-existing urological conditions of bladder outlet obstruction (BOO) or DO, to reflect the instrument's intended target clinical population [12,13]. Female patients were required to have a detrusor pressure at maximum flow ($p_{detQmax}$) of <20 $cmH_20$ and a maximum flow rate ($Q_{max}$) of <15ml/s. Male patients were required to have a bladder contractility index (BCI = $p_{det.Qmax} + 5Q_{max}$) of <100, and a bladder outlet obstruction index (BOOI = $p_{det.Qmax} - 2Q_{max}$) of <40. Five male patients with a BOOI >40 and a BCI <100 were included in the UK concept elicitation study to explore further the range of symptoms reported by patients with DU and co-existing BOO [9].

### *Concept elicitation and item development*

Item generation was based closely on recommendations of the FDA guidance for Industry [10,14] and two reports by the International Society of Pharmacoeconomics and Outcomes Research (ISPOR) on content validity [15,16]. Full description of the qualitative interviews performed in the UK with male and female patients with DU was reported in a previous publication [9]. Decisions surrounding items for inclusion in the initial draft instrument, including their content, language and response options were made using the qualitative evidence and consultation with a multidisciplinary expert panel. Rounds of face to face cognitive interviews with individual patients were scheduled to test and refine the draft items, until all content was considered to be fully understood and interpreted as intended. As patient preference for the length of recall period was inconclusive, two versions were retained to test the effect of two different recall periods during pilot testing; one of 1 week (ICIQ-UAB 1-week) and the other of 24 hours (ICIQ-UAB 24hr).

*Psychometric testing study design*

A non-interventional test-retest pilot study was designed to assess the psychometric properties of the draft instruments in the target population. Patients were identified using their medical records and after written informed consent was given, a study package containing the baseline questionnaires was sent. Patients were asked to complete both ICIQ-UAB versions, and a selection of other PRO instruments of known validity relating to general health, and urinary symptoms and their severity, over a period of 10 days (Table 1). Reliability was evaluated by internal consistency (Cronbach's α), and test-retest reliability by the intraclass correlation coefficient (ICC). Items were considered to have a floor or ceiling effect if more than 20% of respondents answered the lowest or highest of the 5 response options at baseline respectively. The relationship between mean scores and reported symptom severity was investigated using groups stratified by response to the Patient Global Impression of Severity (PGI-S) at baseline (known group validity). Construct validity was assessed using Pearson's correlation coefficient to examine the relationship between the draft ICIQ-UAB versions, and the scores obtained from the other concurrent PRO instruments at baseline.

*Concept elicitation in the US and Japan*

As the ICIQ-UAB is intended for global use, additional concept elicitation interviews were conducted in the US and Japan, to investigate the experiences of patients with UAB in these countries. Interviews were guided by the interview schedule developed during the UK interviews and included open-ended questions, followed by more targeted probing if required. An initial inductive approach [17] to the analysis of the transcripts was followed by the categorisation or 'coding', by existing defined urological symptoms (e.g. 'hesitancy', 'urgency') [2] facilitated by the qualitative analysis software package NVivo v10. The US transcripts were double coded by independent qualitative researchers from the US and the UK and any coding differences reconciled following discussion meetings. For pragmatic reasons, the interviews in Japan were conducted and analysed in Japanese and any coding differences reconciled by review of the patient quotes when translated to English.

*Generation of the developmental ICIQ-UAB*

Following discussion meetings with the clinical experts and the development team, decisions surrounding the modification, reduction, or addition of items were made using evidence from the pilot study and the US and Japanese interviews. The changes proposed to the developmental ICIQ-

UAB were tested using further cognitive interviews with patients that participated in the original UK patient interviews.

**RESULTS**

A total of 143 individual patients were recruited into all sub-studies. A summary of patient demographic and urodynamic characteristics is provided in Table 2.

**Qualitative development of the draft ICIQ-UAB**

Following the UK interviews [9] and subsequent development of a draft instrument, rounds of 3-5 cognitive interviews were conducted with a total of 36 patients. Revisions were made as a result of patient feedback; for example, some DU patients reported voiding infrequently so more discriminatory response options at low frequencies were required. More detailed examples of revisions made are given in the online supplementary material (S1). The resulting ICIQ-UAB 1-week consisted of a total of 31 items (3 medical history items, 20 symptom items, and 8 impact items). For each item the respondent also indicated the level of bother from 0-10, where a score of 0 represents no bother and a score of 10 maximum bother. The ICIQ-UAB 24hr included only the 20 symptom items, as evidence from the cognitive interviews indicated a 24hr recall period was not suitable for the impact items, and the medical history items had their own specific recall periods within the item stem. The layout was based on the template of existing ICIQ modules, previously developed with patient input, and designed to be simple and easy to read, with short items and clear response options [11].

**Pilot study**

54 adult symptomatic patients with a primary urodynamic diagnosis of DU were recruited from 8 sites; 4 in the UK (n=29), 3 in the Netherlands (n=16) and 1 in Germany (n=9). The Dutch and German ICIQ-UAB versions were translated from UK English using linguistic validation methodology to ensure conceptual equivalence [18].

*Missing data*

Missing data at baseline for the ICIQ-UAB 1-week was very low (no responses missing or <3%) for all items and administrations, with the exception of 'clustering of symptoms' (Q23) which had a missing data percentage of 7% at baseline and on retest at day 8. The missing data for ICIQ-UAB 24hr was also very low (no responses missing or <5%).

*Internal consistency and test retest reliability*

Internal consistency was high in both versions (Cronbach's $\alpha \geq 0.85$) (Table 3). ICCs were calculated to be ≥0.88 in all domains and pairs of administrations.

*Item response distribution*

Floor effects in the ICIQ-UAB 1-week at baseline (>20% responses in the lowest option) were found in all three medical history items, many of the symptom items (9/20) and all of the impact items (Table 4). Over half (12/20) of the corresponding symptom items in the ICIQ-UAB 24hr also had floor effects. Only the item measuring 'UTI's over the last month' (Q2) in the ICIQ-UAB 1-week had a

ceiling effect (24% responses in the highest option). In patients with moderate to severe symptoms (n=31) as determined by the PGI-S, the floor effects were reduced to five symptom items in the ICIQ-UAB 1-week and four symptom items in the ICIQ-UAB 24hr (see figure in online supplementary material, S2). Mean bother scores for the symptom items were similar in both versions and ranged from 4.5 to 6.9 (Table 4). The bother scores for the impact and medical history items scored higher for some items; the item for 'planning life around location of toilets', 'self-catheterisation', and 'UTI's over the last month' had the highest mean scores of 7.1, 7.1, and 7.7 respectively.

*Known group validity*

Overall analysis of variance showed a significant difference between the mean scores of the groups of known severity for the symptom items in both versions ($p < 0.05$) (see online supplementary material, S3). The impact items did not quite reach statistical significance ($p=0.052$).

*Convergent and divergent (discriminant) validity*

Pearson correlations of the ICIQ-UAB 1-week and ICIQ-UAB 24hr scores, with the concurrent PROs, were demonstrated with the ICIQ-LUTSqol, MLUTS and FLUTS subscales. Negative correlations were found with the EQ-5D-5L and SF-12v2, where a higher score indicates a better quality of life. The summary correlations are given in the online supplementary material (S4).

**Interviews in the US and Japan**

11 additional concept elicitation interviews were conducted at two hospitals in New York, US. A further 10 interviews were conducted in Nagoya and Fukuoka, Japan. Patients reported largely the same voiding and storage symptoms and similar impact as the UK interviews (see online supplementary material S5-6). However, the US patients reported worry due to the financial burden of their condition (n=3/11), as a result of increased medical expenses and difficulties in securing medical insurance. The emotional impact of their condition could be quite severe: *"I was pretty depressed about it for a while, definitely felt like it affected my manhood".* Several mentioned a detrimental impact on their sex-lives (probably a consequence of medication or surgery) (n=4/11): *"As a single person, losing the ability to have sex – that's a significant issue".* Many of the Japanese patients (n=8/10) described the occurrence of bladder discomfort as a *"dull pain"*, or as a *"heavy"* or *"bloated"* sensation. Personal hygiene was a particular issue for two Japanese patients: "*Because of the slow stream, my trousers get stained…I feel conscious about that."*

**Developmental ICIQ-UAB**

Decisions to modify, remove or add/re-instate items or their response options were made based on the evidence ascertained from all sub-studies, and clinical or patient relevance. For example, the items relating to 'associated bowel symptoms' and 'clustering of symptoms' were removed, due to high floor effects and missing data in the pilot study in addition to low frequency patient reporting and bother in the qualitative phase. The broad emotional impact found in the US interviews was reflected by the more specific examples (e.g. effects on self-esteem, self-confidence) given in the final items. Bladder pain or discomfort was highlighted by the Japanese interviews, and an associated item was added. A dichotomous yes/no item was also added to capture historical self-catheterisation, and the item relating to incontinence was split into two items, to capture and differentiate stress versus urgency incontinence. The additional cognitive interviews conducted with

11 patients confirmed these changes were understood and interpreted as intended. The full item wording and response options along with an example item are given in the online supplementary material (S7-8).

**DISCUSSION**

The ICIQ-UAB is the first PRO measure for the assessment of the symptoms and impact of UAB which is being developed using rigorous methodology, in-line with the FDA Guidance for Industry [10]. To our knowledge, the additional interviews carried out in the US and Japan represents the first qualitative exploration of the patient experience of UAB with patients confirmed to have DU in these countries. The interviews elicited symptoms and impacts that support previous qualitative findings in the UK [9] and the overall content validity of the ICIQ-UAB instrument.

The pilot testing provided encouraging evidence of validity and reliability. Test-retest reliability in stable patients was good (ICC ≥ 0.85 for both versions), providing evidence of the reliability of scores over the test period of 10 days, and that any variability was not due to systematic differences among respondents [19]. The score derived from the tested domains in both versions demonstrated reliability ($\alpha \geq 0.85$) which is over the accepted threshold of ≥0.7 [20] for internal consistency. Many of the items in both versions exhibited floor effects, which can affect the instrument's sensitivity to change [10]. Although these effects were markedly reduced in patients with moderate to severe symptoms, future work on the instrument should look to reduce floor effects in problematic items. The known group validity of the symptom items was supported for both versions, showing that the instrument was sensitive to the reported overall severity of condition. Statistical significance was not quite reached for the impact items possibly due to patient score error, or perhaps a type 2 error as a result of the sample size. Construct validity was demonstrated by the expected convergent and divergent correlations with other PRO measures of known validity. There was no strong evidence from the pilot study to suggest an overall advantage for either recall period version as the item distribution, internal consistency and test-retest reliability was comparable for both versions. Going forward, a shorter recall period is recommended by the FDA guidance for industry to reduce recall bias [10], but this should be carefully considered to match the condition, the PRO domain being measured and views of patients [21].

*Strengths and limitations*

Strengths of the study include the recruitment from multiple sites in Europe, the US, and Japan; the strict adherence to *a priori* urodynamic inclusion criteria, and the first conduct of robust psychometric PRO development methodology in the target population with UAB/DU. The sample characteristics such as the mean age, high PVR and the reported symptoms were comparable to recent publications [13] and are consistent with the recent ICS symptomatic definition of UAB [3].

Professional consensus of urodynamic diagnostic criteria for DU has not yet been achieved for women so a pragmatic approach to the inclusion criteria was taken which reflected parameters used in recent literature. The lower number of females may be a consequence of the referred nature of the sample, for example, there is a greater necessity for diagnostic pressure flow studies performed prior to prostate operations for benign prostatic hyperplasia, to differentiate BOO from DU in men [22]. The ICIQ-UAB is designed as an outcome measure and not as a diagnostic tool for DU/UAB, due to the overlap of reported symptoms with co-existing urological conditions. However, following further

psychometric validation and the development of scoring algorithms it could potentially aid diagnosis when used alongside other non-invasive methods [13]. Underlying DU is known to have a number of possible aetiologies which may be myogenic or neurogenic in origin [23], so further research is required to establish whether any symptomatic differences in presentation may be due to classification of DU by aetiology. In addition, the underlying aetiology of frequently reported symptoms such as nocturia and bladder pain are complex [24], which could affect the responsiveness to an intervention of the associated items. Although the size of the pilot test sample was larger than the recommended minimum size of 40 to obtain reliable estimates of item performance and reliability [25], modifications to the questionnaire as a result of the psychometric testing were deliberately conservative. It is anticipated that the number of items will be reduced using PRO data from larger clinical trials, as well as the evaluation of responsiveness to change and item scoring.

## CONCLUSIONS

The developmental version of the ICIQ-UAB described here is the culmination of a comprehensive set of sub-studies to provide data for its initial validation. The validity, reliability and wider cultural applicability of the ICIQ-UAB among men and women with a confirmed diagnosis of DU are supported by the findings from interviews in the UK, US, and Japan, and European pilot psychometric testing. Following further psychometric testing alongside clinical trials, the instrument is envisaged as an important tool for the monitoring of future treatment strategies for patients with UAB.

## Acknowledgements

The authors would like to thank the patients who participated in this study. This work was funded by Astellas Pharma Europe B.V. who also contributed to the design of the study, preparation, and review of the manuscript.

**References**


1. Chapple CR, Osman NI, Birder L, et al. The Underactive Bladder: A New Clinical Concept? *Eur Urol*. 2015;68(3):351-3.

2. Abrams P, Cardozo L, Fall M, et al. The standardisation of terminology of lower urinary tract function: report from the Standardisation Sub-committee of the International Continence Society. *Neurourol Urodyn*. 2002;21(2):167-178.

3. Wein A, Chapple C. Introduction and Terminology. In: Chapple C, Wein A, Osman N, eds. *Underactive Bladder*. 1st Edition. Switzerland: Springer; 2017:ix-xiii.

4. Osman N, Hillary C, Chapple C. Epidemiology of Underactive Bladder. In: Chapple C, Wein A, Osman N, eds. *Underactive Bladder*. 1st Edition. Switzerland: Springer; 2017:25-31.

5. Jeong SJ, Kim HJ, Lee YJ, et al. Prevalence and Clinical Features of Detrusor Underactivity among Elderly with Lower Urinary Tract Symptoms: A Comparison between Men and Women. *Korean J Urol*. 2012;53(5):342-348.

6. Hoag N, Gani J. Underactive Bladder: Clinical Features, Urodynamic Parameters, and Treatment. *Int Neurourol J*. 2015;19(3):185-189.

7. Abarbanel J, Marcus E-L. Impaired detrusor contractility in community-dwelling elderly presenting with lower urinary tract symptoms. *Urology*. 2007;69(3):436-440.

8. Coyne KS, Wein AJ, Tubaro A, et al. The burden of lower urinary tract symptoms: evaluating the effect of LUTS on health-related quality of life, anxiety and depression: EpiLUTS. *BJU Int*. 2009;103 Suppl 3:4-11.

9. Uren AD, Cotterill N, Harding C, et al. Qualitative Exploration of the Patient Experience of Underactive Bladder. *Eur Urol*. 2017;72(3):402-407.

10. FDA. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. December 2009.

11. Abrams P, Avery K, Gardener N, Donovan J, ICIQ Advisory Board. The International Consultation on Incontinence Modular Questionnaire: www.iciq.net. *J Urol*. 2006;175(3 Pt 1):1063-1066; discussion 1066.

12. Abrams P. Bladder outlet obstruction index, bladder contractility index and bladder voiding efficiency: three simple indices to define bladder voiding function. *BJU Int*. 1999;84(1):14-15.

13. Gammie A, Kaper M, Dorrepaal C, Kos T, Abrams P. Signs and Symptoms of Detrusor Underactivity: An Analysis of Clinical Presentation and Urodynamic Tests From a Large Group of Patients Undergoing Pressure Flow Studies. *Eur Urol*. 2016;69(2):361-369.

14. FDA. Guidance for Industry and FDA Staff: Qualification for Drug Development Tools. January 2014.

15. Patrick DL, Burke LB, Gwaltney CJ, et al. Content Validity—Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 1—Eliciting Concepts for a New PRO Instrument. *Value Health*. 2011;14(8):967-977.

16. Patrick DL, Burke LB, Gwaltney CJ, et al. Content Validity—Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 2—Assessing Respondent Understanding. *Value Health*. 2011;14(8):978-988.

17. Thomas DR. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *Am J Eval*. 2006;27(2):237-246.

18. Wild D, Grove A, Martin M, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health*. 2005;8(2):94-104.

19. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Third edition. Oxford University Press; 2004.

20. Nunnally J, Bernstein I. The assessment of reliability. In: *Psychometric Theory*. Third ed. McGraw-Hill Inc; 1994:248-292.

21. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013;22(8):1889-1905.

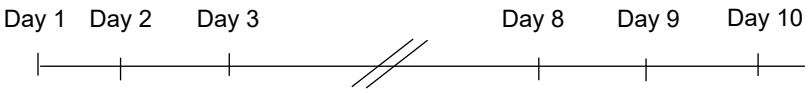
22. Cho MC, Ha SB, Park J, et al. Impact of Detrusor Underactivity on Surgical Outcomes of Laser Prostatectomy: Comparison in Serial 12-Month Follow-Up Outcomes Between Potassium-Titanyl-Phosphate Photoselective Vaporization of the Prostate (PVP) and Holmium Laser Enucleation of the Prostate (HoLEP). *Urology*. 2016;91:158-166.

23. Dewulf K, Abraham N, Lamb LE, et al. Addressing challenges in underactive bladder: recommendations and insights from the Congress on Underactive Bladder (CURE-UAB). *Int Urol Nephrol*. 2017;49(5):777-785.

24. Drake MJ. Should nocturia not be called a lower urinary tract symptom? *Eur Urol*. 2015;67(2):289-290.

25. Hertzog MA. Considerations in determining sample size for pilot studies. *Res Nurs Health*. 2008;31(2):180-191.

Table 1. Psychometric testing study flow chart and concurrent PRO measures used.




| | | | | | | | |
|---|---|---|---|---|---|---|---|
| DHIF | x | | | | | | |
| **ICIQ-UAB 1-week** | x | | | | x | | |
| **ICIQ-UAB 24h** | x | x | x | | x | x | x |
| ICIQ-MLUTS /FLUTS | x | | | | | | |
| ICIQ LUTSqol | x | | | | | | |
| SF-12 v2 | x | | | | | | |
| EQ-5D-5L | x | | | | | | |
| PGI-S | x | | | | x | | |
| PGI-C | | | | | x | | |

**Abbreviations and concurrent PROs used:** DHIF, demographic health and information form; EQ-5D-5L, EuroQol-5 dimensions (Five domains for capturing health-related quality of life); ICIQ, International Consultation on Incontinence Questionnaire; FLUTS, female lower urinary tract symptoms (LUTS and bother in females 13 item questionnaire); LUTSqol, lower urinary tract symptoms quality of life (impact of LUTS on health related quality of life 22 item questionnaire); MLUTS, male lower urinary tract symptoms (LUTS and bother in males 13 item questionnaire); PGI-C, patient global impression of change (single rating of the change of status of a subject's symptoms on a 7 point scale); PGI-S, patient global impression of severity (single rating of severity of subject's condition on a 4 point scale); SF-12 v2, short form-12 item version 2 (questionnaire for health related quality of life over the last 4 weeks).

**Note:** Approximately half of the participants completed the ICIQ- UAB 1-week first at baseline (day 1) followed by the concurrent PRO measures and the ICIQ-UAB 24hr at the end. For the other half of the participants the order of the UAB PROs was reversed to eliminate possible order effects of administration.

Table 2. Summary of patient characteristics for each of the sub studies.

| **Patient characteristic** | **Total dataset*** | **UK CE interviews** | **UK cognitive interviews** | **Japanese CE interviews** | **US CE interviews** | **European Pilot study** |
|---|---|---|---|---|---|---|
| Total number of patients n | 143 | 44 | 36 | 10 | 11 | 54 |
| Males n (%) | 109 (76) | 29 (66) | 27 (75) | 7 (70) | 10 (91) | 42 (78) |
| Females n (%) | 34 (24) | 15 (34) | 9 (25) | 3 (30) | 1 (9) | 12 (22) |
| Age [years] median (IQR) | 66 (17) | 69 (18) | 68 (13) | 68 (17) | 69 (32) | 65 (17) |
| ISC current or historical n (%) | 50 (35) | 23 (52) | 11 (31) | 4 (40) | 5 (46) | 12 (22) |
| PVR >0ml n (%) | 113 (79) | 36 (82) | 30 (83) | 9 (90) | 9 (82) | 41 (76) |
| PVR [ml] median (IQR) | 190 (257) | 183 (402) | 130 (290) | 164 (168) | 235 (850) | 206 (271) |
| BCI (males only) median (IQR) | 75 (29) | 72 (27) | 72 (16) | 68 (27) | 90 (55) | 80 (27) |
| BOOI (males only) median (IQR) | 19 (20) | 18 (24) | 20 (19) | 21 (14) | 0 (15) | 23 (19) |
| $P_{detQmax}$ [$cmH_20$] median (IQR) | 29 (22) | 25 (23) | 31 (21) | 34 (19) | 18 (11) | 31 (21) |
| $Q_{max}$ [ml/s] median (IQR) | 7.9 (5) | 8 (5) | 8 (4) | 5.5 (5) | 14 (11) | 8 (6) |
| BVE (%) median (IQR) | 67 (59) | 56 (59) | 71 (59) | 53 (57) | 43 (62) | 73 (54) |

*Although 36 cognitive interviews were performed, 12 patients had previously participated in the CE interviews, so to avoid duplication the data for these patients are only included once in this column. **Abbreviations:** CE, concept elicitation; ISC, intermittent self-catheterisation; IQR, interquartile range; $P_{detQmax}$, detrusor pressure at maximum flow; $Q_{max}$, maximum flow rate; BCI, bladder contractility index calculated by BCI=$p_{det}Q_{max}$+5$Q_{max}$ ; BOOI, bladder outlet obstruction index calculated by $p_{det}Q_{max}$-2$Q_{max}$ ; PVR, post void residual; BVE, bladder voiding efficiency = voided volume/total bladder capacity×100.

Table 3. Test-retest reliability and internal consistency of the ICIQ-UAB 1-week and ICIQ-UAB 24hr.

| | **Test-retest reliability ICC (95% CI) (n=42)*** | **Internal consistency Cronbach's α (n=54) ǂ** |
|---|---|---|
| ICIQ-UAB 1-week (Days 1-8) | | |
| Symptom items | 0.90 (0.82-0.94) | 0.85 |
| Impact items | 0.90 (0.83-0.95) | 0.85 |
| ICIQ-UAB 24hr | | |
| Days 1-2 | 0.88 (0.78-0.93) | 0.86 |
| Days 2-3 | 0.92 (0.86-0.96) | |
| Days 8- 9 | 0.93 (0.88-0.96) | |
| Days 9-10 | 0.94 (0.88-0.97) | |

**Abbreviations**: ICC, Intraclass correlation coefficient; CI, Confidence intervals. ICIQ-UAB 1-week, underactive bladder patient-reported outcome (1-week recall); ICIQ-UAB 24hr, underactive bladder patient-reported outcome (24-hour recall);

* Stable patients as determined by the Patient Global Impression of Change (PGI-C) tool at day 8.

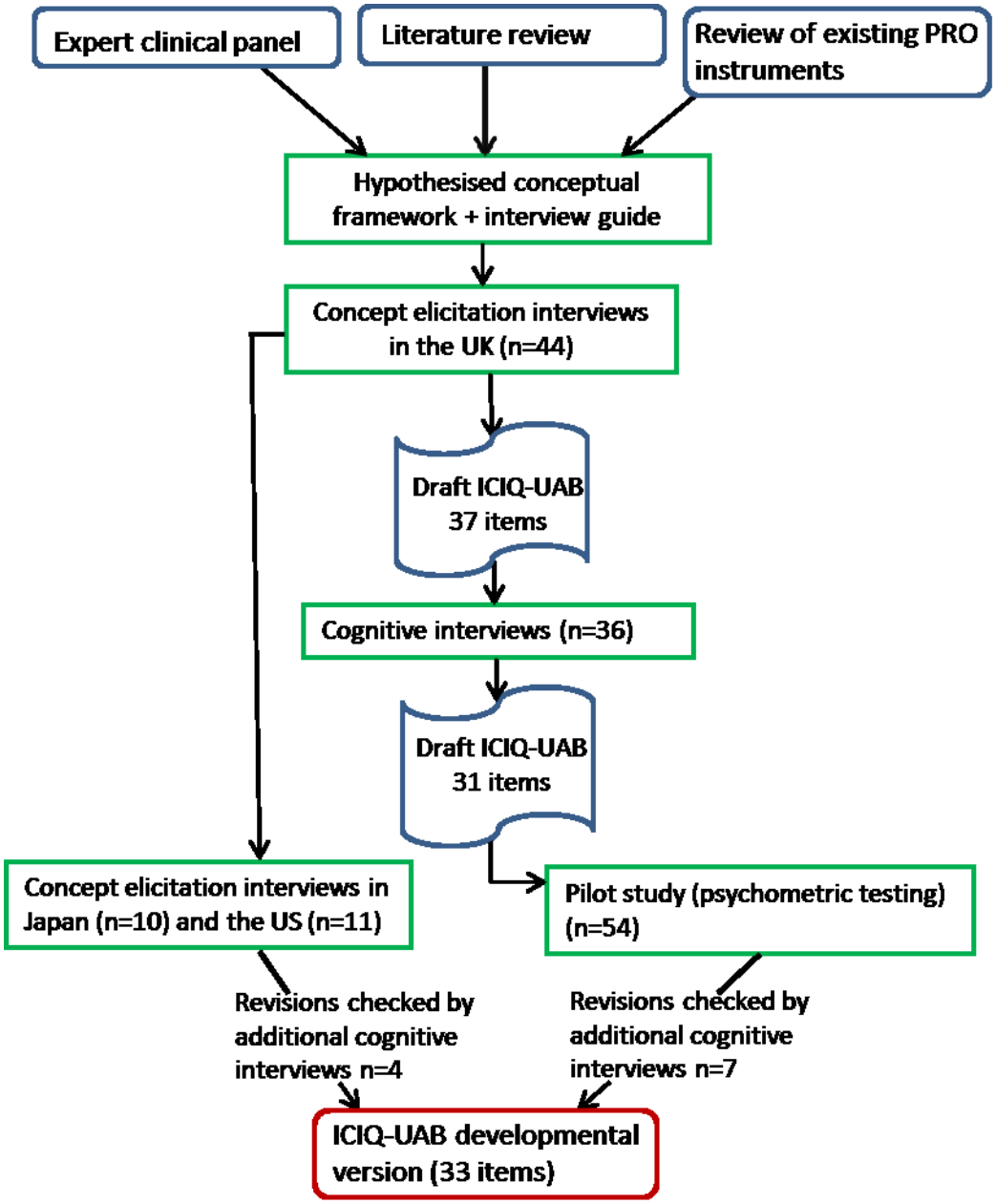
ǂ Full analysis set (n=54) at day 1.

Table 4. The percentage of responses given for the lowest and highest response option with average bother score for each recall period version and item, at baseline.

| Item (ICIQ-UAB 1-week/ICIQ-UAB 24hr) | % of responses in lowest response option (n=54) | | % of responses in highest response option (n=54) | | Bother item mean score* (SD) | |
|---|---|---|---|---|---|---|
| | ICIQ-UAB 1-week | ICIQ-UAB 24hrs | ICIQ-UAB 1-week | ICIQ-UAB 24hrs | ICIQ-UAB 1-week | ICIQ-UAB 24hrs |
| Acute retention (Q1) | 83‡ | | 4 | | 6.4 (3.4) | |
| UTIs over last month (Q2) | 59‡ | | 24‡ | | 7.7 (2.5) | |
| Self-catheterisation (Q3) | 80‡ | | 2 | | 7.1 (2.3) | |
| Hesitancy (Q4/Q1) | 9 | 13 | 9 | 11 | 4.6 (2.8) | 5.0 (3.1) |
| Need to concentrate to void (Q5/Q2) | 28‡ | 20‡ | 11 | 9 | 5.2 (2.7) | 5.3 (3.0) |
| Small volume per void (Q6/Q3) | 15 | 22‡ | 4 | 4 | 5.6 (2.7) | 5.8 (3.0) |
| Post-micturition dribble (Q7/Q4) | 50‡ | 52‡ | 2 | 4 | 6.6 (2.4) | 6.6 (2.5) |
| Incontinence (stress and urgency) (Q8/Q5) | 70‡ | 67‡ | 0 | 0 | 6.9 (2.6) | 6.3 (2.7) |
| Need to immediately re-void (Q9/Q6) | 28‡ | 37‡ | 2 | 2 | 6.3 (2.3) | 6.6 (2.5) |
| Sensation of incomplete emptying (Q10/Q7) | 13 | 19 | 11 | 11 | 5.8 (2.8) | 6.0 (3.0) |
| Intermittency (Q11/Q8) | 15 | 19 | 4 | 9 | 5.7 (2.6) | 5.4 (3.0) |
| Straining (to begin void) (Q12/Q9) | 24‡ | 30‡ | 4 | 11 | 5.8 (2.8) | 5.7 (3.2) |
| Straining (to end void) (Q13/Q10) | 9 | 20‡ | 11 | 13 | 5.8 (2.7) | 6.3 (2.9) |
| Slow stream (Q14/Q11) | 15 | 15 | 2 | 2 | 5.5 (2.8) | 5.2 (3.0) |
| Urgency (Q15/Q12) | 19 | 26‡ | 0 | 2 | 6.1 (2.6) | 6.3 (2.4) |
| Nocturia and/or nocturnal voids (Q16/Q13) | 9 | 11 | 11 | 13 | 5.8 (3.0) | 5.5 (3.1) |
| Daytime urinary frequency (Q17/Q14) | 7 | 7 | 4 | 2 | 4.9 (3.1) | 4.9 (3.3) |
| Reduced sensation of bladder fullness (Q18/Q15) | 24‡ | 20‡ | 2 | 6 | 4.9 (3.1) | 5.1 (3.0) |
| Waiting in bathroom after voiding (Q19/Q16) | 24‡ | 26‡ | 11 | 4 | 5.6 (2.8) | 5.5 (2.6) |
| Length of time in bathroom (Q20/Q17) | 15 | 15 | 2 | 2 | 5.0 (2.9) | 5.1 (3.3) |
| Temporarily unable to pass urine (Q21/Q18) | 61‡ | 67‡ | 0 | 0 | 6.0 (3.2) | 5.8 (3.1) |
| Associated bowel symptoms (Q22/Q19)** | 78‡ | 82‡ | 4 | 4 | 6.5 (2.7) | 6.8 (3.0) |
| Clustering of symptoms (Q23/Q20)*** | n/a | n/a | n/a | n/a | 4.5 (3.6) | 4.5 (3.7) |
| Planning life around location of toilets (Q24) | 41‡ | | 6 | | 7.1 (2.2) | |
| Social life (Q25) | 50‡ | | 7 | | 6.4 (2.5) | |
| Nocturia/nocturnal voids impact (Q26) | 41‡ | | 9 | | 6.7 (2.5) | |
| Physical activities (Q27) | 46‡ | | 4 | | 6.1 (2.5) | |
| Feelings about self (Q28) | 46‡ | | 11 | | 6.3 (3.1) | |
| Embarrassment (Q29) | 43‡ | | 4 | | 6.6 (2.7) | |
| Fluid intake (Q30) | 30‡ | | 9 | | 5.8 (2.6) | |
| Overall impact (Q31) | n/a | | n/a | | 5.1 (3.3) | |

‡ These items showed a floor or ceiling effect (≥20% of responses were in lowest or highest category). *Mean scores for bother were calculated using only those who experienced the symptom. **The first part of this item had a yes/no response option so the floor effect was calculated using the percentage who answered 'no' to this item. ***This item had nominal response options so a floor effect was not applicable.

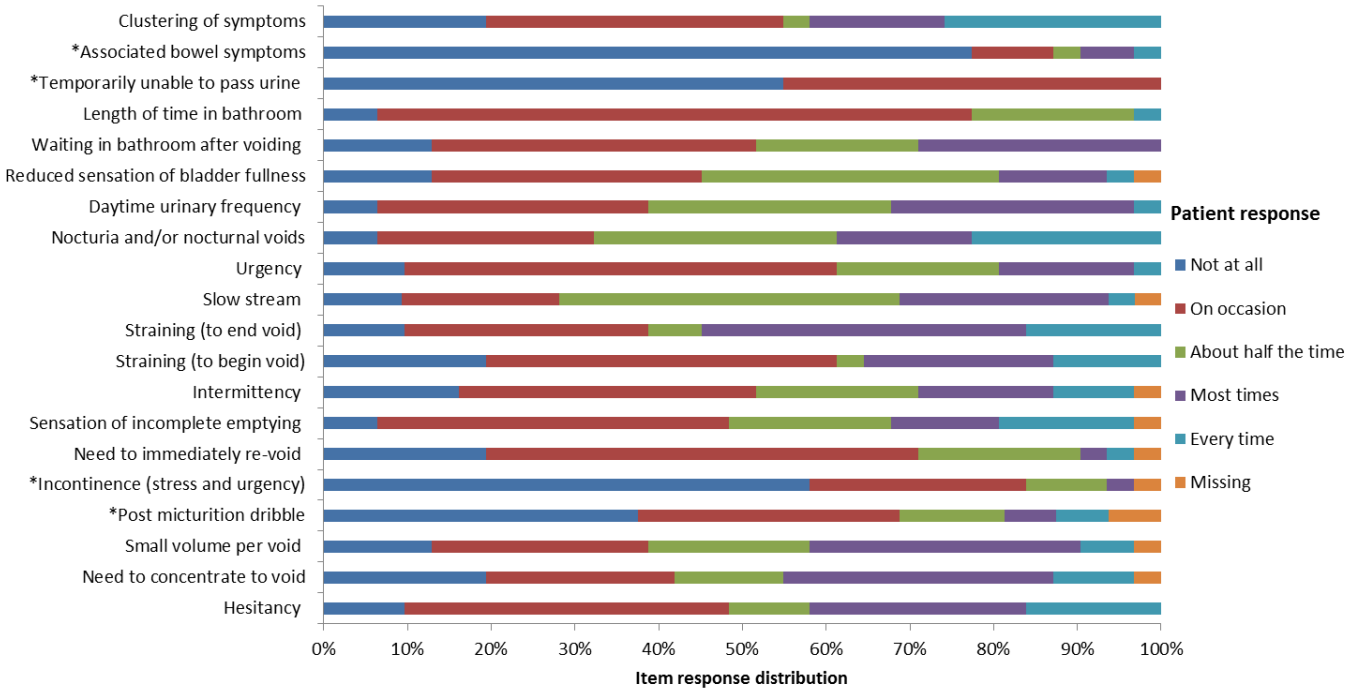
Figure 1. The development of the ICIQ-UAB.



S1. Examples of revisions based on feedback from the patients in the cognitive interviews.

| **Questionnaire component** | **Rationale** | **Patient and clinical panel quotes** | **Revision** |
|---|---|---|---|
| **Recall period**<br><br>A 24 hour and 1 week recall period for symptom items and 4 weeks for the impact items | Many of the patients reported that they had not seen the recall period. For symptom items there was no consensus but a longer recall period was usually preferred for the impact questions. | P6"I was thinking of a day, a normal day"<br><br>P22:"I felt better when I have to think about it over the last month to give a more reliable picture of my condition" | The recall period was put as a reminder at the top of each item box to encourage respondents to adhere to the recall period provided. Two versions with a 24 hour and 1 week for symptom items were retained for further psychometric testing. A longer recall period for impact items. |
| **Item clarity**<br><br>**Original item:** How often do you feel like you were not able to pass what you might consider a satisfactory amount of urine? | Some patients found this item unclear or interpreted this as an item asking about a sensation of incomplete emptying. | P4: "You don't get any relief anymore"<br><br>Clinical panel: "A bit of a confusing question" | Item reworded to improve patient understanding and interpretation.<br><br>**Item:** How often were you only able to pass a small volume of urine? |
| **Item wording**<br><br>**Original item**: Do you have to strain or squeeze to urinate? | The wording of 'strain' was preferred to 'squeeze' by patients in the context of starting to pass urine. | P2: "I don't know about squeeze I mean straining it's the force to try and you know, to activate…" | Replaced with:<br><br>**Item:** How often did you strain to start your urinations? |
| **Item relevance**<br><br>**Item:** A pictorial representation of a slow stream which represented the projection of flow when standing up. | Although well understood by men this was not applicable to women as they do not stand up to urinate. A sex independent item was preferred. | P22: "I don't see how that [item] is relevant to women" | Replaced with:<br><br>**Item:** On average, would you say that that the strength of flow of your urinary stream was… (Normal (not reduced), a little reduced, reduced, very reduced, extremely reduced) |
| **Response option**<br><br>Item: How often do you pass urine during the day? (1-6 times, 7-8 times, 9-10 times, 11-12 times, 13 or more times) | Some DU patients void quite infrequently so more discriminatory options were required at lower voiding frequencies. | P8: "I'd rerate it there... I'd probably go for three"<br><br>Clinical panel: "Some UAB sufferers void very infrequently" | Response options revised:<br><br>Item: During the day, how many times did you urinate, on average? (1-3 times, 4-6 times, 7-9 times, 10-12 times, 13 or more times) |

S2. Distribution of responses at baseline for ICIQ-UAB 24hr in patients with moderate to severe symptoms (n=31). *Items which met the criteria for a floor effect (>20% responses in the lowest option).



S3. Mean score of ICIQ-UAB 1-week and ICIQ-UAB 24hr domains when stratified into known groups determined by the responses to the PGI-S at baseline.

| **Known groups (PGI-S)** | **N*** | **ICIQ-UAB 1-week** | | | | **ICIQ-UAB 24hr** | |
|---|---|---|---|---|---|---|---|
| | | **Symptom items mean score (SD)** | **P value** | **Impact items mean score (SD)** | **P value** | **Symptom items mean score (SD)** | **P value** |
| No-symptoms | 7 | 26.4 (12.3) | 0.002 | 14.3 (25.5) | 0.052 | 21.6 (11.1) | 0.001 |
| Mild | 11 | 26.8 (12.1) | | 20.1 (17.4) | | 22.6 (11.0) | |
| Moderate | 27 | 41.8 (12.4) | | 35.0 (22.0) | | 36.7 (12.1) | |
| Severe | 4 | 39.9 (8.7) | | 39.3 (17.7) | | 37.9 (8.3) | |

**Note:** Scores are transformed and range from 0-100 with higher scores representing higher severity of symptoms or impact. *PGI-S missing data n=5.

S4. Pearson's correlations between ICIQ-UAB 1-week and ICIQ-UAB 24hr subscales, and concurrent PROs at baseline.

| Concurrent PRO measure domain | ICIQ-UAB 1-week | | ICIQ-UAB 24hr |
|---|---|---|---|
| | Symptom items | Impact items | Symptom items |
| ICIQ LUTSqol (Social Limitation) | 0.67 | 0.89 | 0.67 |
| ICIQ LUTSqol (Severity) | 0.37 | 0.50 | 0.29 |
| ICIQ LUTSqol (Role Limitation) | 0.55 | 0.69 | 0.50 |
| ICIQ-MLUTS (Voiding domain) | 0.74 | 0.34 | 0.76 |
| ICIQ-MLUTS (Incontinence domain) | 0.55 | 0.58 | 0.45 |
| ICIQ-FLUTS (Filling domain) | 0.52 | 0.68 | 0.52 |
| ICIQ-FLUTS (Voiding domain) | 0.77 | 0.27 | 0.75 |
| ICIQ-FLUTS (Incontinence domain) | 0.32 | 0.07 | 0.35 |
| EQ-5D-5L VAS | -0.29 | -0.40 | -0.35 |
| SF-12v2 Physical Component Score | -0.28 | -0.42 | -0.32 |
| SF-12v2 Mental Component Score | -0.33 | -0.54 | -0.35 |

**Note**: A correlation of 0.1-0.3 is considered weak, 0.3-0.5 is moderate and ≥ 0.5 is strong [29]. On ICIQ questionnaires, higher score indicates severe symptom or greater impact on QoL; on SF-12v2, higher score indicates better health; on ICIQ-UAB 1-week, higher score indicates severe symptoms and impact; on EQ-5D-5L VAS, zero indicates the worse health state and 100 the best health state.

**Abbreviations:** EQ-5D-5L, EuroQol-5 dimensions (Five domains for capturing health-related quality of life); ICIQ, International Consultation on Incontinence Questionnaire; FLUTS, female lower urinary tract symptoms (LUTS and bother in females 13 item questionnaire); LUTSqol, lower urinary tract symptoms quality of life (impact of LUTS on health related quality of life 22 item questionnaire); MLUTS, male lower urinary tract symptoms (LUTS and bother in males 13 item questionnaire); PGI-C, patient global impression of change (single rating of the change of status of a subject's symptoms on a 7 point scale); PGI-S, patient global impression of severity (single rating of severity of subject's condition on a 4 point scale); SF-12 v2, short form-12 item version 2 (questionnaire for health related quality of life over the last 4 weeks); VAS, visual analog scale.

S5. The symptoms or sign concepts reported during the UK, US and Japanese concept elicitation interviews.

| **Symptom/sign (concept)** | **Japanese n=10** | **US n=11** | **UK n=44** | **Patients who reported concept (n=65) %** |
|---|---|---|---|---|
| Nocturia | ✔ | ✔ | ✔ | 75 |
| Slow stream | ✔ | ✔ | ✔ | 68 |
| Daytime frequency of urination | ✔ | ✔ | ✔ | 58 |
| Hesitancy | ✔ | ✔ | ✔ | 55 |
| Straining | ✔ | ✔ | ✔ | 55 |
| Urgency | ✔ | ✔ | ✔ | 52 |
| Sensation of incomplete emptying | ✔ | ✔ | ✔ | 52 |
| Urinary incontinence | ✔ | ✔ | ✔ | 45 |
| Intermittent self-catheterisation | ✓ | ✔ | ✔ | 43 |
| Post micturition dribble | ✔ | ✔ | ✔ | 43 |
| Intermittent stream | ✔ | ✔ | ✔ | 40 |
| Urinary tract infection | ✓ | ✔ | ✔ | 35 |
| Need to immediately re-void | ✓ | ✔ | ✔ | 34 |
| Bladder discomfort/lower urinary tract pain | ✔ | ✔ | ✔ | 34 |
| Urinations of small volume | ✔ | ✔ | ✔ | 29 |
| Temporarily unable to pass urine | ✓ | ✔ | ✔ | 25 |
| Clustering of urinations | ✓ | ✔ | ✔ | 20 |
| Splitting/spraying | ✔ | **X** | ✔ | 18 |
| Reduced sensation of bladder fullness | ✔ | ✔ | ✔ | 18 |
| Associated bowel symptoms | **X** | ✔ | ✔ | 17 |
| Acute retention | **X** | ✔ | ✔ | 17 |
| Bloated sensation | ✔ | **X** | ✓ | 5 |
| Urinations of long duration | ✔ | ✓ | ✓ | 2 |
| Coldness (bladder sensation) | ✔ | **X** | **X** | **2** |

✔ The symptom was described by patients in the qualitative interviews.

✓ The symptom was described by patients but coded (or recorded) differently by the country study researcher (e.g. 'temporarily unable to pass urine' was coded as 'hesitancy' (severe) in the Japanese study; urinary tract infections and use of intermittent self catheterisation were recorded on the case report form in the Japanese study).

**X** The symptom was not described by patients in the qualitative interviews.

S6. The impact concepts reported during the UK, US and Japanese concept elicitation interviews.

| **Impact (concept)** | **Japanese n=10** | **US n=11** | **UK n=44** | **Patients who reported concept (n=65) %** |
|---|---|---|---|---|
| Planning activities around location of toilets | ✔ | ✔ | ✔ | 65 |
| Manages fluid intake | ✔ | ✔ | ✔ | 42 |
| Feelings about self/emotional | ✔ | ✔ | ✔ | 38 |
| Sleep disturbance/Tired | ✔ | ✔ | ✔ | 37 |
| Embarrassment | ✔ | ✔ | ✔ | 35 |
| Family and friends | ✔ | ✔ | ✔ | 28 |
| Social life | ✓ | ✔ | ✔ | 17 |
| Affects physical activity | ✔ | ✔ | ✔ | 17 |
| Sex life | **X** | ✔ | ✔ | 12 |
| Financial | **X** | ✔ | **X** | **5** |
| Frustration | ✔ | ✓ | ✓ | 3 |
| Hygiene | ✔ | **X** | ✓ | 3 |
| Affects clothing | ✔ | **X** | **X** | **2** |

✔ The impact was described by patients in the qualitative interviews.

✓ The impact was described by patients but coded differently by the country study researcher (e.g. 'Frustration' was coded as 'feelings about self/emotional' in the US and UK studies).

**X** The impact was not described by patients in the qualitative interviews.

S7. An example of an item in the developmental ICIQ-UAB.

**Over the <u>LAST 24 HOURS…</u>**

**14a. How often did you strain to <u>start</u> passing urine?**

| | | |
|---|---|---|
| not at all | ☐ | 0 |
| occasionally | ☐ | 1 |
| about half the time | ☐ | 2 |
| most of the time | ☐ | 3 |
| every time | ☐ | 4 |

**b. How much does this bother you?**
*Please circle a number between 0 (not at all) and 10 (a great deal)*

**0** 1 2 3 4 5 6 7 8 9 **10**
not at all a great deal

S8. The developmental ICIQ-UAB 33-item list with response options.

| | |
|---|---|
| 1. Have you ever been unable to pass urine at all and had to go to hospital to have a catheter tube inserted to drain the bladder? | no, once, twice, three or more times |
| 2. Over the last 12 months, have you had a urinary infection for which you took medication? | no, unsure, once, twice, three or more times |
| 3. Have you ever self-catheterised? | yes, no |
| 4. Over the last week, how often did you self-catheterise? | not at all, less than once a day (1-6 times), 1-2 times a day, 3-4 times a day, 5 or more times a day |
| Over the LAST 24 HOURS… | |
| 5. When ready to pass urine, was there a delay before the urine flow started? | not at all, occasionally, about half the time, most of the time, every time |
| 6. When ready to pass urine, did you feel you had to concentrate in order to start passing urine? | not at all, occasionally, about half the time, most of the time, every time |
| 7. How often were you only able to pass a small volume of urine? | not at all, occasionally, about half the time, most of the time, every time |
| 8. How often did a few drops leak out into your underwear shortly after you had finished passing urine and had dressed yourself? | not at all, occasionally, about half the time, most of the time, every time |
| 9. How often did you leak urine before you could get to the toilet? | not at all, occasionally, about half the time, most of the time, every time |
| 10. How often did you leak urine when physically active, or when you coughed or sneezed? | not at all, occasionally, about half the time, most of the time, every time |
| 11. After passing urine, how often did you have to return to the bathroom to pass urine again, within a short space of time (e.g. within 15 minutes)? | not at all, occasionally, about half the time, most of the time, every time |
| 12. Soon after passing urine, how often did you have a sensation that your bladder was not completely empty? | not at all, occasionally, about half the time, most of the time, every time |
| 13. When passing urine, how often did the flow stop and start? | not at all, occasionally, about half the time, most of the time, every time |
| 14. How often did you strain to start passing urine? | not at all, occasionally, about half the time, most of the time, every time |
| 15. How often did you strain to maintain your flow when passing urine? | not at all, occasionally, about half the time, most of the time, every time |
| 16. How often did you strain towards the end of passing urine, to try and empty your bladder? | not at all, occasionally, about half the time, most of the time, every time |
| 17. On average, would you say that the strength of flow of your urinary stream was… | normal (not reduced), a little reduced, reduced, very reduced, extremely reduced |
| 18. How often did you experience a sudden or strong need to pass urine which you were unable to ignore, and had to rush to the bathroom? | not at all, occasionally, about half the time, most of the time, every time |
| 19. During the night, how many times did you have to get up to pass urine? | not at all, once, twice, three times, four or more times |
| 20. During the day, how many times did you pass urine? | (1-3 times, 4-7 times, 8-9 times, 10-11 times, 12 or more times) |
| 21. Did you find it difficult to tell when your bladder | not at all, occasionally, about half the time, |

| | |
|---|---|
| was full? | most of the time, every time |
| 22. How often did you stay a bit longer in the bathroom after passing urine, to make sure your bladder was as empty as possible? | not at all, occasionally, about half the time, most of the time, every time |
| 23. What was the longest time that you needed to spend in the bathroom trying to empty your bladder? | 1-2 minutes, 3-5 minutes, 6-10 minutes, 11-15 minutes, more than 15 minutes |
| 24. How often did you go to the toilet to pass urine but were unable to urinate at all, so had to return to the bathroom to try again later? | not at all, occasionally, about half the time, most of the time |
| 25. How often did you feel pain or discomfort in your bladder as it filled? | not at all, occasionally, about half the time, most of the time, every time |
| Over the LAST WEEK… | |
| 26. How often did you make plans around the location of toilets (e.g. shopping, social outings, travelling, holidays)? | not at all, occasionally, sometimes, most of the time, all of the time |
| 27. How often did you feel that your urinary symptoms interfered with your normal daily activities (e.g. social life, work outside the home)? | not at all, occasionally, sometimes, most of the time, all of the time |
| 28. How often did you feel that getting up at night to pass urine affected your day to day life? | not at all, occasionally, sometimes, most of the time, all of the time |
| 29. How often did your urinary symptoms prevent you from getting the amount of sleep you needed? | not at all, occasionally, sometimes, most of the time, all of the time |
| 30. How often did you feel that your urinary symptoms affected your physical activities (e.g. walking, swimming, sport)? | not at all, occasionally, sometimes, most of the time, all of the time |
| 31. Did your urinary symptoms affect the way you feel about yourself (e.g. embarrassment, self-confidence, self-esteem)? | not at all, occasionally, sometimes, most of the time, all of the time |
| 32. Did your urinary symptoms cause you to be careful about how much or the type of fluid you drink? | not at all, occasionally, sometimes, most of the time, all of the time |
| 33. Overall, how much would you say your urinary symptoms interfered with your everyday life? | not at all, occasionally, sometimes, most of the time, all of the time |