Urinary stones and Intervention Quality of Life (USIQoL): Development and validation of a new core universal patient reported outcome measure for urinary calculi

(Abstract word count: 272, Paper word count: 2516)

ABSTRACT

Background:

Urolithiasis has significant impact on patients' health-related quality of life (HRQoL).

Objective:

To develop a core patient-reported outcome measure (PROM), using modern psychometric methods, to quantify the impact of urolithiasis and different treatments.

Design, Setting, and Participants:

Adult patients with urinary calculi, attending urology departments, covering all index categories and treatment spectrum, participated during different phases. The pilot instrument was created from the potential items (phase 1 and 2) within conceptual framework. It was pretested (Phase 3) followed by psychometric evaluation in two parts (Phases 4 and 5).

Outcome Measurements and Statistical Analysis:

Validity and reliability assessments of the new PROM were performed with Rasch Measurement Theory (RMT) (RUMM 2030) and traditional analyses.

Results:

In total, 683 patients [median age 51 years (range 18-92)] participated during different phases. The initial 60 item draft (5 scales) was completed by 212 patients (Phase 4). A revised 25 item draft was produced after removal of unstable items. In field test 2, the revised version was evaluated by 369 patients producing the final USIQoL (15 items) with summated logit scores. It includes three scales, pain with physical health (6), psychosocial health (7) and work performance (2). The lower scores indicate better outcomes. It was found to be reliable ($r \geq 0.8$), internally consistent ($\alpha \geq 0.7$), to have good construct validity (good hypothesised correlations, $r > 0.3$) with satisfactory sensitivity to change ($p < 0.01$). All scales demonstrated uni-dimensionality with good item fit and person separation indices.

Limitation:

USIQoL has been developed in the English language within UK population.

Conclusions:

The USIQoL is a short unidimensional, valid and reliable PROM to assess the HRQoL impact of urinary calculi and treatments based on RMT. It is expected to serve as a core PROM across entire spectrum of urolithiasis.

Patient Summary

Kidney stone is a common condition with various treatment options. The condition and treatments have significant impact on patient's quality of life. This can be measured objectively using valid and reliable patient reported outcome measure (PROM) developed using modern methods. We have developed a PROM that provides helpful and accurate measurement useful for all stakeholders.

(Abstract word count: 272, Paper word count: 2516)

1. Introduction

Urolithiasis is a common condition with the prevalence of 2-3% amongst the general population with 50% patients more likely to form further stones in five years [1]. The reported prevalence rates range from 7-13% in North America, 5- 9% in Europe and 1-5% in Asia [2]. The disease resulted in 550,000 emergency room visits in the US in 2009 and over 30,800 hospital admissions in England in a year [3-4. Stone patients miss an average of 47.9 hours of work/year with additional hours lost due to ambulatory care visits [5]. There are different options to manage urinary calculi with expectant, medical or interventional treatments [6], which can be multi-staged and carry different risks and success rates. Temporary interventions such as indwelling ureteric stents add to the patient burden [7]. Urolithiasis and its treatment(s) have an adverse effect on health-related quality of life (HRQoL) and can compromise all areas of patient functioning [8-10].

A patient reported outcome measure (PROM) is a report on patient's health condition that comes directly from the patient. [11]. In addition, to their use in randomised controlled trials to assess treatment effectiveness, there is growing interest in their use in the routine HRQoL monitoring and medical audits [12]. A PROM, could improve the evidence base, as long as the measure is appropriate and accord with international standards [13]. American Urological Association (AUA) guidelines state that treatment decisions about urinary calculi should incorporate patient preferences that are influenced by the HRQoL impact [14].

Attempts have been made to measure HRQoL of patients with urolithiasis. Generic measures have been used for this, but often fail to elaborate on the clinically relevant domains [8]. In the recent past, PROMs specific to urolithiasis targeted at different sub-populations have been developed [15-17]. It is now recognised that modern psychometric methods based on (RMT) should be integral to the development of such measures [18,19].

Our hypothesis was subjective quality of life impact of the stone disease and interventions can be measured in an objective fashion with the use of a valid and reliable patient reported outcome measure developed using modern methodology. Our aim was to develop a core PROM, incorporating RMT, to evaluate the impact of entire

spectrum of upper tract urinary calculi in a uniform way and facilitate cross comparison of HRQoL impact of urolithiasis as well as interventions.

2. Materials and Methods

We followed international PROM guidelines for the development and validation of the Urinary Stones and Intervention Quality of Life (USIQoL) measure that would also conform to the COSMIN (COnsensus-based Standards for the selection of health Measurement Instruments) checklist [19]. The multicentre developmental process comprised of 5 stages (Ethical approval: 17/WA/0195, no. 138478, 217163). Adult patients with urolithiasis covering all index stone categories, representative of the routine practices, were invited to participate. These included patients with renal or ureteric stones with or without treatment(s). The key steps are outlined in Figure 1.

2.1 Phases 1- 3: This work, with patient interviews, involving many stakeholders, produced working conceptual framework and initial long draft of the questionnaire (8, 20). The revised draft, after pretesting, was administered in the field test 1.

Phase 4: Field test 1: This was undertaken to construct USIQoL scales and perform a preliminary psychometric evaluation in a large sample to select most appropriate items.

Phase 5: field test 2: This was undertaken to comprehensively evaluate the shortened version, from phase 4, to produce final draft. Patients also completed existing generic questionnaires and Ureteric Stent Symptoms Questionnaire (USSQ) with indwelling stent (21).

2.2: Sample size considerations and Statistical Analysis

The rule of thumb sample size recommendations for traditional (10 subjects/item of the largest subscale (18 items in the long draft, n=180) and Rasch analysis (n= min 200 and max 400/500, for four/five class intervals) were followed, for all assessments, during phases 4 and 5

A combination of traditional and RMT assessments were conducted using sophisticated mathematical measurement model [22]. SPSS 25 software was used to perform traditional (for exampleSpearman correlations) analysis. Rasch analysis

(polytomous extended response category, partial credit model) was performed using RUMM 2030 software.

2.2 Rasch Analysis (See Glossary)

Phase 4: It assessed different properties of the USIQoL like item and person locations, item fit (fit residuals and X^2 statistics), Person Separation Index (PSI), response categories and local dependence (23). The removal of mis-fitting items was conducted in an iterative manner with the removal of single item at a time followed by re-run of the analyses.

Phase 5: In addition to the phase 4, we assessed 1) Differential item functioning (DIF) for the traits: a) age (four groups), b) sex, c) stone site (kidney/ureter), d) type of interventions, e) presence or absence of symptoms and f) history of previous stones, 2) Smith's test of uni-dimensionality (24) and 3) optimal scale structure and logit-based scoring.

2.3 Traditional analysis (internal consistency and validity)

Phase 4: Inter-item and corrected item total correlations were calculated. Correlations between scales (EuroQoL EQ-5D, SF12, (Hospital Anxiety and Depression scale [HADS] or Work Productivity and Activity Impairment Questionnaire [WPAI] {expected 0.3-0.5}) were assessed for criterion validity (25-28).

Phase 5: In addition to the phase 4, we conducted tests of reliability (test-retest, patients with stable disease completing the USIQoL, twice, 24-72 hours apart) and validity (convergent) including both within and between scales testing and responsiveness to change (subgroup completing the USIQoL before and after interventional treatments with 4-16 weeks interval).

3. Results

3.1 Item generation (Figure 1)

A total of 62 from 77 invited patients (mean age 51 years) and 30 family members participated in the Phase 1 and 2, generating 106 themes and 10 broad headings. These were mapped to conceptual framework with removal of duplications to create

item sets [8, 20]. A five-point rating scale (‘not at all’ to ‘a lot’) was selected for the initial draft

3.2 Pre-testing

Forty patients evaluated USIQoL, with minor changes to the items providing preliminary evidence for its face content validity and clinical suitability. The review by the clinicians confirmed its completeness. The revised versions with 60 items including treatment items - were drafted for the first-field test. It evaluated pain using different formats, (frequency of mild to unbearable pain, intensity of worst, day to day as well as average pain [10 items]), physical and social health including sex life (18 items), psychological health (6 items), work performance (8 items) and travel/holiday issues (3 items). Fourteen items addressed additional problems including treatments and help from the healthcare and family members and 1 global health.

3.3 Field-test one - item reduction and scale development

During the first field-test, 212/250 patients completed the questionnaires (Table 1). We evaluated psychometric properties considering it to be a single scale and seven sub-scale formats.

Rasch analysis: This demonstrated important features of USIQoL, including limitations, requiring modifications. All scales indicated ‘good to excellent’ reliability (PSI 0.62-0.89, Table 2). However, almost all scales had over 60% of the items with disordered thresholds (difficulty distinguishing between responses ‘quite a bit’ and ‘very much’) necessitating change from five to 4 or 2 response categories. Each scale had items with significant fit residuals (12%-60%), and residual correlations (50-90%), indicating item redundancy.

Traditional analysis: This showed USIQoL to be reliable and valid measure of impact of stones on different domains. Reliability was excellent [alpha: total scale (0.9), subscales (0.6-0.9)]. The corrected item total (0.3-0.8) and inter-item (0.4-0.9) correlations were satisfactory. Preliminary analyses of criterion validity were as expected (correlations with generic measures, range 0.3-0.8) demonstrating satisfactory early item level validity.

Using an iterative approach, the mis-fitting and redundant items were removed to generate the revised versions for Stone disease and interventions (20 and 24 items respectively). It included five scales of pain, social health (5 items each), physical, psychological health (4 each) and work (2) with 4 treatment items ready for field test 2.

3.4 Field-test two

In total**,** 369 of 390 patients participated in the phase 5 [409 observations, 61 patients completing >1 (pre- and post-treatment) questionnaires] with 24 of 30 patients completing test-retest study [Table 1].

Rasch analysis: This demonstrated that most of the items in the scales mapped out continua of increasing bother (Table 2). The scales located items in a clinically sensible order with good sample match. Deviations from model expectations were marginal.–Items excluded were pain (life interference, average and mild pain), social (sex, social life, and holiday), psychological (worry about kidney failing), and treatment (diet and device). The two treatment items (medication, water intake) were combined with the social scale. This transformed the USIQoL into a final 15-item measure.

Revised scaling: Items had superior fits when the 5-scale structure was changed to 3-scale, combining pain and physical health domains (PPH 6 items), psychological and social health domains (PSH 7 items) and work domain (2 items). Figure 2 demonstrates satisfactory item-threshold distribution maps of sub-scales.

DIF and Uni-dimensionality: We evaluated all 15 questions, and 3 scales, against different patient sub-populations (Appendix B). This confirmed good performance across traits. All 3 scales were uni-dimensional.

Traditional analysis: This confirmed all three USIQoL scales to be reliable and valid measures for assessment of important domains across patient groups. Corrected item-total - and inter-item correlations provided support that items within scales measured a common underlying construct with good reliability. Test-retest correlations were excellent (0.81-0.92) indicating good scale stability (Table 3).

Criterion validity was tested extensively and hypothesised correlations between scores from USIQoL scales and existing generic and domain specific measures were consistent. We showed that there was very good correlation with the relevant domains

between USIQoL and USSQ scales. The USIQoL was responsive to change as shown by significant positive effect in all scale scores after intervention.

Final USIQoL measure and scoring (Appendix B): The final USIQoL (3 scales and 15 items) is intended for self-administration where patients rate the amount of bother attributed on a 4-point (1=not at all, 2=a little, 3=quite a bit or 4=a lot). The disease and intervention versions are exactly similar and differ only in the title time frame (since your 'current stone problems' or 'current or most recent stone treatment') to make it psychometrically valid. Scale scores are generated by summing items and transferring to a 0-100 (logit) scale with high scores indicating greater patient bother.

4. Discussion

The recurrent nature and ensuing interventions for urolithiasis can result in a cumulative negative HRQoL impact. This can be assessed using PROMs, however, carry measurement challenges. Generic measures fail to capture this impact comprehensively. Hence, urolithiasis specific PROMs were developed recently. The Wisconsin Stone Quality of Life (WISQoL -28 items) measure was the first one to be developed to assess the impact of stable urolithiasis and medical therapies [15-17]. It has undergone linguistic validations with wide applications in different studies.

It is well recognised that measures that comply with modern psychometric methods based on item response theory (RMT) to be of higher quality (19). In this respect, development of the recent PROMs had a focus on specific subgroups and involved only traditional methods. These don't cover key criteria from the COSMIN guidelines (content development, sample size, use of RMT, uni-dimensionality etc).

The new USIQoL is a first PROM to capture the HRQoL impact of urolithiasis (acute and chronic) and interventions. It is developed using a combination of classical and RM theory with very few such measures in urology. In the Rasch model, the probability of a specified response (right/wrong answer) is modelled as a function of person and item parameters. It is a unique approach of mathematical modelling based upon a latent trait where item values are calibrated, and person abilities are measured on a shared continuum that accounts for the latent trait. It provides an internally valid measure that is independent of the particular sample with the findings for the sample extrapolating to its population and measure clinically meaningful differences [29, 30].

The final 15-item selection in USIQoL was based on the appraisals of the analyses against clinical relevance and measurement criteria. The psychometric evaluation showed that all three scales satisfy criteria for acceptability, validity and reliability. The logit scoring for each scale offers different scores allowing clearer identification of the impact across different domains. This would help future comparative studies and sample size calculations.

The results from traditional validity assessments alone, suggested that the long draft of the USIQoL satisfied most of the criteria, until RMT demonstrated many targeting problems (disordered responses, item redundancies). This highlighted the value of RMT to conduct item-level analyses that guide precise item selection and rectify problems with scales. It demonstrated that our five-stage mixed methods approach to be important due to complex assessments.

4.1 Strengths and limitations

Many aspects important to different stakeholders were considered during conceptual framework and subsequent steps. Apart from the construction of necessary items and scales based on the key themes, we carefully evaluated if there was a need for separate instruments for renal and ureteric stones as well as disease and interventions. We also looked at demonstrable applicability of the PROM and uniformity of the performance across entire disease spectrum. Our work indicated that, the QoL surrounding different sites, disease and treatments are interlinked and separate measures can pose psychometric difficulties. The USIQoL development phases demonstrated that formulation of integrated single PROM gave a better model of item fit and performed well across patient, disease and intervention groups making USIQoL appropriate core instrument.

All three scales of the USIQoL demonstrated very good performance with proven uni-dimensionality. It is observed that pain along with physical symptoms, which drive most of the clinical assessments, have more visible impact. The domain of pain, the most complex to assess, was tested extensively before finalising its appropriate format for inclusion. Similarly, issues regarding work are important to all stakeholders. On the other hand, psychosocial scale is likely to be a good indicator of the issues not evaluated routinely and the longer-term impact of the condition, which could drive treatment choices. The USIQoL captures all these dimensions well with the results

quantified using modern psychometric techniques. The USIQoL can also help reliable combined HRQoL evaluation for stent subgroup.

There are certain limitations of the study and future work would help to address these. USIQoL was developed and validated in the English-speaking population of the UK and its wider application would need linguistic and cross-cultural validation. Its application with existing measures, such as WISQoL, could help capture a broader picture. Its further applications, including daily practices, will investigate scale sensitivity to develop clinically relevant thresholds. There is a scope for adaptations to undertake economic appraisals and compare emergent treatments and service evaluations which would guide patient centric care.

5. Summary

In conclusion, the USIQoL is a new three scale 15-item, single page, self-report instrument that measures the HRQoL impact of stone disease and interventions. It has been developed using modern psychometric methods (RMT). It is fit for valid and reliable comparisons at the micro-level (patients) and meso-level (treatment groups, institutions). We expect USIQoL to serve as a core PROM for studies looking at, and comparing effectiveness of the treatments, observational strategies, quality of care as well as an adjunct to the medical audits. This is expected to improve the evidence base and help improved patient communication and shared decision making.

Acknowledgements:

The Urological Foundation (TUF) for part funding. We thank Prof. K Sarica, Dr. F Sangueodolce (EAU), Mr. A Dickinson and BAUS Endourology Committee Members, Prof. M Monga. and Dr. Srilingam (AUA). We also thank Prof. S. Salek for help with the study design.

Abbreviations and Glossary of Terms

HRQoL: health related quality of life

COSMIN: COnsensus-based Standards for the selection of health Measurement INstruments

PROMs: Generic

1. **EQ-5D** -3L: EuroQol questionnaire Descriptive system for health-related quality of life states in adults, consisting of five dimensions. It is a preference based HRQoL measure also used for economic appraisals (to calculate quality adjusted life years, QALY)
2. **SF12** - Medical Outcomes Study (short form 120 self-reported outcome measure assessing the impact of health on an individual's everyday life. Also used for preference based utility (economic) assessments.

Disease and Intervention specific

1. **USIQoL**: Urinary Stones and Intervention Quality of Life questionnaire: Urinary Stones (upper tract) disease and intervention specific QoL measure
2. **WISQoL**: Wisconsin Stone Quality of Life questionnaire (WISQOL): Kidney stone disease specific QoL measure
3. **USSQ**: Ureteral Stent Symptoms Questionnaire - Intervention (ureteral stent) specific measure to assess impact of ureteral stents on QoL
4. **HADS** - Hospital Anxiety and Depression Scale - To measure anxiety and depression in a general medical population of patients
5. **WPAI**: Work productivity and Activity Impairment Scale - to measure impairments in work

Rasch Measurement Analysis terminology
(item = question, trait = patient/disease characteristics)

1. **RMT:** Rasch measurement Theory
2. **Logit range** – For information on scale to sample targeting by measuring the match between range of HRQoL (domains) measured by the USIQoL and the range of HRQoL in the patient sample
3. Targeting and **Person Separation Index (PSI)**: measures reliability of the scale. A questionnaire is perfectly targeted if the mean of the person is the same as the mean of the items on the shared metric. PSI represents the extent to which items distinguish between distinct levels of disease-specific bother
4. **Item-fits**: a) The chi squared statistics to measure that the central property of item invariance (the hierarchical ordering of the items) does not vary across the trait measured b) fit residuals: the differences between the observed and expected data for each person and item
5. **Ordered thresholds**: Consistent use of the scale that corresponds to the evidence that the response categories represent increasing levels of the construct being measured (the correct ordering of the response categories is reflected in successive thresholds)
6. **Residual correlation**: The extent to which each item is independent of the others (helps to remove redundant questions)

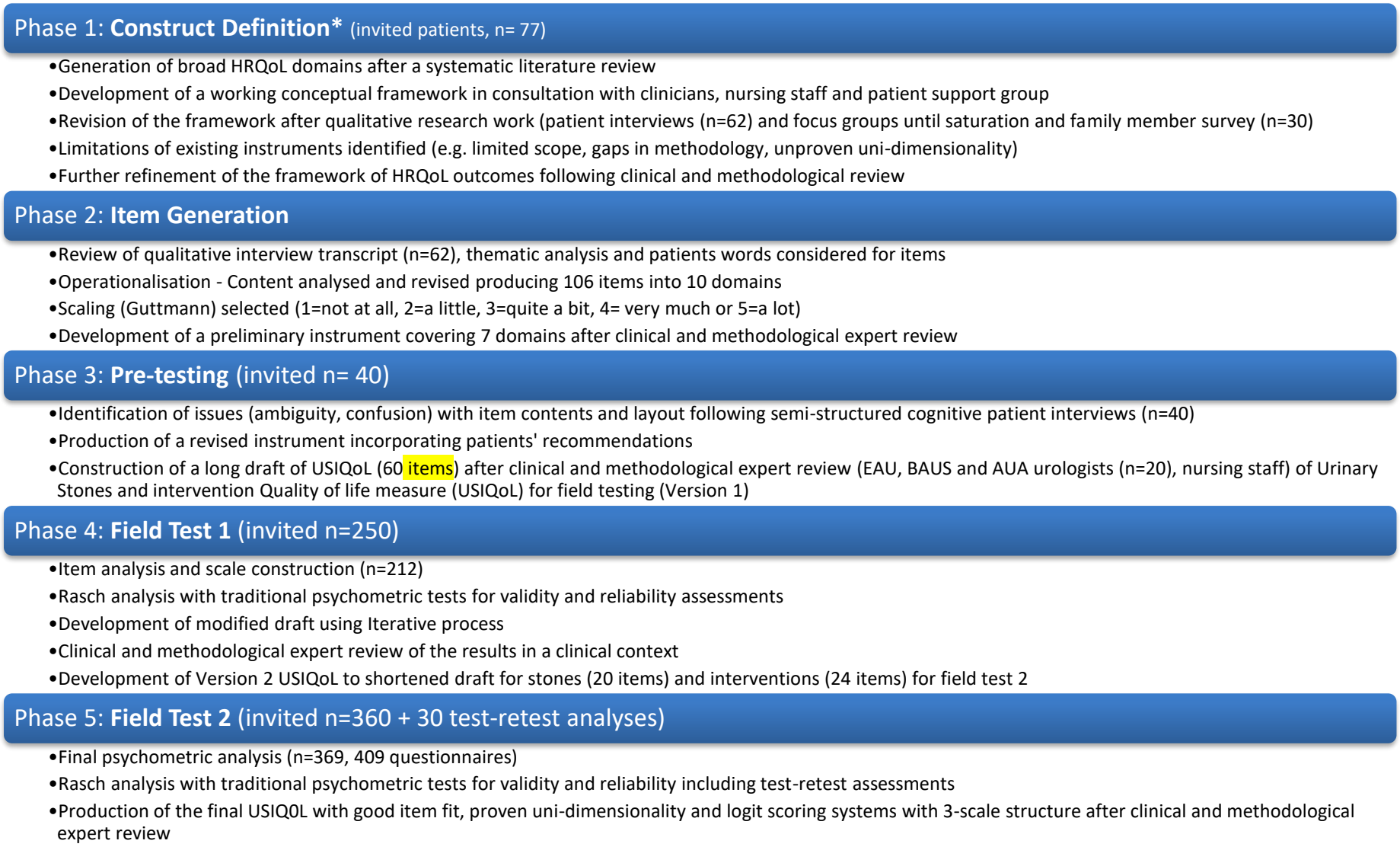
7. **Uni-dimensionality**: To determine if there are any other identifiable dimensions in the data after the main "Rasch dimension" has been taken into account.
8. **DIF**: Differential item functioning: Evaluates the extent to which different groups within the sample (e.g. age, site of stone, type of intervention)

References:

1. Pearle MS, Calhoun EA, Curhan G: Urologic diseases in America project: urolithiasis. J. Urol 2005; 173: 848.
2. Sorokin I, Mamoulakis C, Miyazawa K, Rodgers A, Talati J, Lotan Y: Epidemiology of stone disease across the world : World J. Urol; 2017; 35(9): 1301-20.
3. Hospital Episodes Statistics Data. Vol 2014. National Health Service2014.
4. Bultitude M, Rees J. Management of renal colic. BMJ. 2012;345:e5499.
5. Saigal CS, Joyce G, Timilsina AR et al: Direct and indirect costs of nephrolithiasis in an employed population: opportunity for disease management? Kidney Int 2005;68:1808-14.
6. Türk C, Neisius A, Petrik A, Seitz C, Skolarikos A, Thomas K. European Urological Association urolithiasis guidelines http://uroweb.org/guideline/urolithiasis/ 2020.
7. Joshi HB, Stainthorpe A, McDonagh RP et al: Indwelling ureteral stents: Evaluation of symptoms, Quality of Life and Utility. J. Urol: 169, 1065-1069.
8. Raja A, Hekmati Z, Joshi HB. How Do Urinary Calculi Influence Health-Related Quality of Life and Patient Treatment Preference: A Systematic Review. J Endourol. 2016;30:727-743.
9. Penniston KL, Nakada SY. Treatment expectations and health-related quality of life in stone formers. Curr Opin Urol. 2016;26:50-55.
10. Patel N, Brown RD, Sarkissian C, De S, Monga M. Quality of life and urolithiasis: the patient - reported outcomes measurement information system (PROMIS). Int Braz J Urol. 2017;43:880-886.
11. U.S Dept of Health and Human Service Food and Drug administration: Guidance for Industry: Patient reported outcome measures – use in medical product development to support labelling claims. Silver Spring: U.S. dept of health and Human Service: FDA; 2009.
12. Dept of Health. Equity and Excellence: liberating the NHS. London Dept of health 2010.
13. US FDA and Scientific Advisory Committee of the Medical Outcomes Trust: Assessing health status and quality of life instruments: attributes and review criteria. Qual. Res. 2002, 11:193-205.
14. Assimos D, Krambeck A, Miller NL, et al. Surgical Management of Stones: American Urological Association/Endourological Society Guideline, PART I. J Urol. 2016;196:1153-1160.
15. Penniston KL, Nakada SY. Development of an instrument to assess the health related quality of life of kidney stone formers. J Urol. 2013;189:921-930.
16. Tran MGB, Sut MK, Collie J. et al: Development of a disease-specific ureteral calculus patient reported outcome measurement instrument J. Endourol; 32:6:548-558.

17. Ragab M, Baldin N, Collie J et al: Qualitative exploration of the renal stone patients' experience and development of the renal stone-specific patient-reported outcome measure BJUI 2020; 125: 123-32.
18. Lord FM. Applications of item response theory to practical testing problems Hillsdale NJ: Erlbaum 1980, Rasch G: Probabilistic models for some intelligence and attainment tests, Copenhagen, 1960, expanded edition Chicago, University of Chicago Press; 1980
19. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159-70.
20. Raja A, Wood F, Joshi, HB. The impact of urinary stone disease and their treatment on patients' quality of life: a qualitative study: Urolithiasis. 2020; 48(3): 227–234.
21. Rasch G. Probabilistic models for some intelligence and attainment test. Chicago: University of Chicago 1960.
22. Smith EV Jr. Metric development and score reporting in Rasch measurement: J. Appl. Meas. 2000;1:303-26.
23. EuroQoL - a new facility for the measurement of health related quality of life. The EuroQol group. health Policy 1990; 16: 199.
24. Gandek B, Ware JE, Aaronson NK et al. Cross-validation of Item Selection and Scoring for the SF-12 Health Survey in Nine Countries: Results From the IQOLA Project. International Quality of Life Assessment J Clin Epidemiol 1998 Nov; 51(11):1171-8.
25. Zigmond AS, Snaith RP. The Hospital Anxiety And Depression Scale, Acta Psychiatrica Scandinavica, 1983: 67, 361-370.
26. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics. 1993;4:353–365.
27. Joshi HB, Newns N, Stainthorpe A, MacDonagh RP, Keeley FX, Jr, Timoney AG. Ureteral stent symptom questionnaire: development and validation of a multidimensional quality of life measure. J Urol. 2003;169:1060-1064.
28. Smith EVJr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas, 2002: 3(2), 205-231.
29. Granger C. Rasch Analysis is Important to Understand and Use for Measurement, American Educational Resaerch Association, www.rasch.org/rmt, Rasch Measurement Transactions, 2007, 21:3 p. 1122-3.
30. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods: Health Technol Assess. 2009;13(12):1-200.

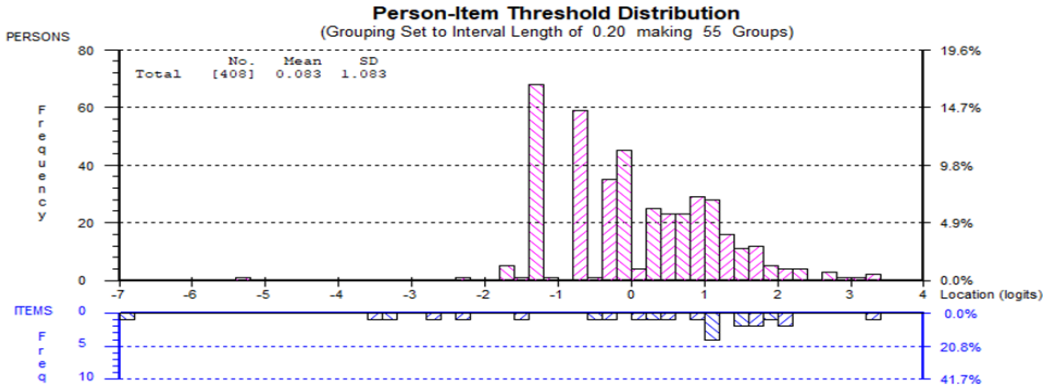
Figure 1: Steps in the Development (Phases 1-2) and Evaluation (Phases 3-5) of USIQoL:



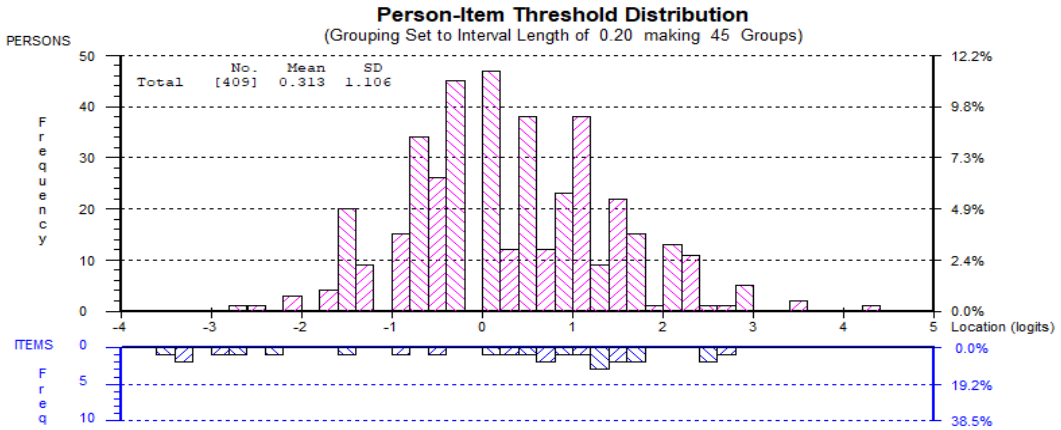
*__Construct Definition__: PROM development underpinned by the theory (conceptual base) and consideration of the target population. Establish need of stones and intervention specific PRO: Map construct from existing instruments to conceptual framework and follow latest methodology

Figure 2: Person – Item Thrshold Distribution map

A) Pain and Physical Health Scale (PPH)



B) Psychosocial Health Scale (PSH)



C) Work

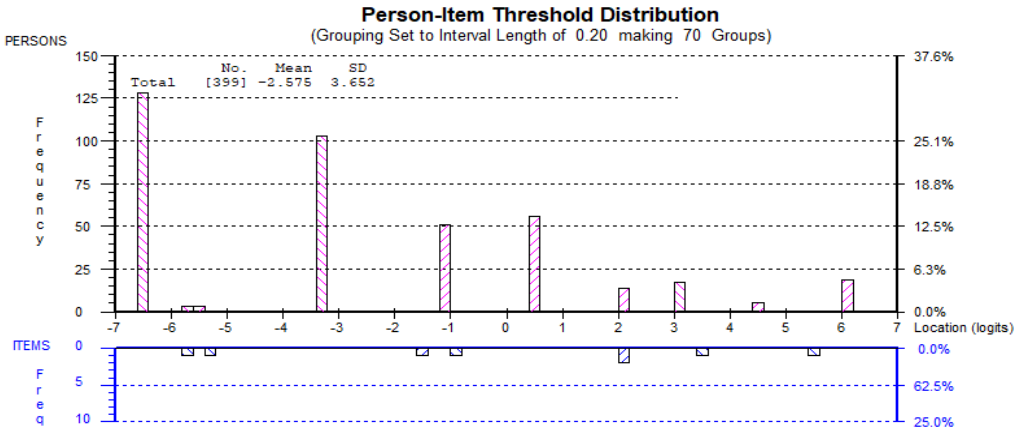


Table 1: Participant Characteristics

Characteristics	**Field Test 1 (n=212 of 250)**	**Field Test 2 (n=345 of 360*)**
Age: Mean/Median	52.5/52 yr.	53.8/56 yr.
Age (Range)	19-89 yr.	18-90 yr.
16-40 yrs.	64 (30%)	98 (29%)
41-64 yrs.	92 (43%)	166 (48%)
65-80 yrs.	50 (24%)	74 (21%)
>80 yrs.	6 (3%)	7 (2%)
Sex: M:F [missing]	135:77 (64%:36%)	241:103(69%:31%) [0.3]
Bothersome pain: Y/N[N.R.]	127 (65%) /73 (35%) /12 [5.8%]	204(60%)/138(40%)/3 [1%]
Site of stone: Kidney[N.R.]	147 (68%)/3 [1.4%]	225 (63%) /5[1.4%]
Site of stone: Ureter	62 (32%)	115 (35%)
Previous stones: Y/N [N.R.]	116/84 (56%/44%) /12 [5.8%]	176/162 (51%/48%) / 7 [2%]
Paid employment: Y/N [N.R.]	125/81/ [6]	230/115
Current Treatment(s)		
Medical(Metabolic disorder)	13 (6%)	28 (8.1%)
SWL	73 (34.4%)	122 (35.3%)
Surgical interventions (URS/PCNL)	37 (18.3%)	79 (22.8%)
Observation (with/without short term medical treatment tamsulosin, analgesics)	88 (41.3%)	117 (33.8%)
Stent in situ	16 (7.5%)	33 (9.6%)

(*Test-retest data not included, N.R. – not recorded)

Table 2: Results of Psychometric testing: Rasch analysis: Field Test 1 (Phase 4, Long draft, n=212) [single scale and important subscales] and Field test 2 (Phase 5, n=345 with 409 observations)

USIQoL - Scale (no. of Items)	*Items (Mean)		*Persons (Mean)		**Person Separation index	***Fit Statistics	***Items with X^2 Prob>0.001	§Dis-ordered Thresholds	§§Item-residual correlation	§§§Uni-dimensionality (PCP)
	locations (Logit range)	Fit residual	locations (Logit range)	Fit residual		(FR outside+/- 2.5)			'r' score range >+0.3	't' tests <5% (Y/N)
Field Test 1 (n=212)										
Total (59)	0.0	-0.08	-0.27	-0.45	0.88	14	10	55	56	N
Pain (10)	0.0	-0.94	-0.36	-0.69	0.84	4	2	6	7	N
Physical + Social (18)	0.0	-0.46	-0.08	-0.60	0.89	4	2	16	12	N
Psychological (6)	0.0	-0.88	-0.34	-0.98	0.89	1	1	0	3	N
Work (8)	0.0	-0.85	-1.10	-0.62	0.81	7	5	8	6	N
Travel (3)	0.0	-0.87	-0.66	-0.50	0.62	0	2	3	2	N
Field Test 2 (n=345) – Final Draft USIQoL										
Pain and Physical Health (6)	0.00	-0.88	0.08	-0.61	0.72	0(-2.0 to +1.3)	0	0	0(-0.3 to+0.07)	Y (4.35%)
Psycho-Social Health (7)	0.00	-0.58	0.31	-0.55	0.70	0(-1.0 to +0.3)	0	0	0(-0.4 to+0.06)	Y (3.5%)
Work (2)	0.00	-1.14	-2.58	-1.05	0.83	0(-1.3 to -0.9)	0	0	0 (-0.98)	Y (1.65%)

*Logit range – For information on scale to sample targeting (match between range of HRQoL measured by the USIQoL and the range of HRQoL in the patient sample)

**Person Separation Index (PSI): to measure the reliability of the scale (0.7 being adequate)

***Item-fits measured by a) fit residuals (expected to lie between a mean of 0 and SD±2.5) and b) chi squared statistics (should be under Bonferroni corrected significance level)

§Disordered Thresholds: Response categories not working as intended (Measured by item response curves and threshold maps)

§§Residual correlation: The extent to which each item is independent of the others (should be <0.3 above mean)

§§§Uni-dimensionality: Smith's test of uni-dimensionality within scales. (This identifies if the person estimates derived from the most diverse subsets of items are significantly different using principal component analysis. If the proportion, or the lower bound of the 95% confidence interval of significant ($p<0.05$) t-tests, is less than 5% it indicates uni-dimensionality)

Table 3: Summary of USIQoL Traditional Psychometric analysis: Phase 5 (Field test -2)

	USIQoL Scales (no. of items)		
Test Criteria (Observations)	Pain and Physical Health PPH (6 items)	Psycho-Social Health PSH (revised 7 item scale, n=156)	Work Performance (2 items)
Cronbach's alpha	0.82	0.75	0.94
Inter Item Correlations (Range)	0.29-0.56	0.11-0.59	0.89
Item Total Correlations (range)	0.51-0.62	0.0.35 -0.58	NA
Test retest (n=24)	0.91	0.83	0.80
Construct Validity (Correlation coefficients)			
EQ-5D-3L (n=346)			
EQ-5D-3L utility - Total	-0.41	-0.36	0.02
EQ-5D-3L-Pain/Discomfort	0.49	0.39	0.12
EQ-5D-3L-Mobility	0.26	0.24	-0.11
EQ-5D-3L- Usual Activities	0.40	0.42	0.02
EQ5D – Anxiety/Depression	0.28	0.51	0.12
EQ5D-Thermometer	-0.39	-0.45	0.01
SF 12			
SF12 – Physical health (PCS, n= 296)	-0.53	-0.54	-0.00
SF12 – Mental health (MCS, n= 302)	-0.37	-0.46	-0.15
WPAI			
WPAI (n=67)	0.46	0.62	0.7
HADS			
HADS – Anxiety (n=166)	0.47	0.52	0.08
HADS – Depression (n=163)	0.54	0.52	-0.00
USSQ			
USSQ – Pain (n=14)	0.71	0.30	NA
USSQ – Urinary Symptoms (n=14)	0. 84	0.56	NA
USSQ – Gen. Health (n=14)	0.62	0.87	NA
*Pre-Post treatment scale: effect size (n=57)	0.6	0.123	0.35

WPAI- Work Productivity and Activity Impairment Scale

HADS - Hospital Anxiety and Depression Scale

USSQ – Ureteral Stent Symptoms Questionnaire

*(We assessed the responsiveness of the USIQoL by calculating effect sizes comparing 'pre' and 'post' intervention scale scores. We expected these to be positive confirming post treatment improvement leading to reduced bother but did not hypothesise a magnitude given relatively smaller sample and first application of the scales)

Appendix A
Urinary Stones and Intervention Quality of Life Measure©
The USI-QoL – Intervention

We are interested in knowing how your **quality of life** has been affected since your current, or most recent, treatment(s) for urinary stones. N/A Not applicable

Since your current or most recent stone treatment(s) and due to your stone treatment(s), **how much do you suffer with**	Not at all	A little	Quite a bit	A lot	N/A
Q1. Severe to unbearable pain?					
Q2. Pain triggered by physical activity?					
Q3. The feeling you need to pass urine urgently?					
Q4. Symptoms of a urinary tract infection (e.g. running temperature, feeling unwell and pain while passing urine)?					
Q5. Decreased or lack of appetite?					
Q6. Low energy?					
					PPH Total □□
Since your most recent stone treatment(s), **how much have you**	Not at all	A little	Quite a bit	A lot	N/A
Q7. Had difficulty sleeping?					
Q8. Felt depressed?					
Since your most recent stone treatment(s), **with regards to the future, how much are you worried about:**					
Q9. More symptoms from your stones in the future?					
Since your most recent stone treatment(s), **have your symptoms made you reluctant about:**	Not at all /A little		Quite a bit/A lot		N/A
Q10. Making a long journey?					
Since your most recent stone treatment(s), **how much have you had to visit the following, due to your symptoms:**	Not at all	A little	Quite a bit	A lot	N/A
Q11. GP or hospital during normal working hours?					
Since your most recent stone treatment(s), **how much have you found yourself having problems with:**					
Q12. Having to take medication (painkillers, preventative treatment etc.)?					

Q13. Increasing your water intake?					
					PSH Total □□

Work

Please mark 'Not applicable' (N/A) if currently not working (paid employment).					
Since your most recent stone treatment(s), **with regard to your job, how much:**	Not at all	A little	Quite a bit	A lot	N/A
Q14. Have you needed to take time off work?					
Q15. Has your stone disease interfered with your ability to do your job?					
					WorkTotal □□

Appendix B: Diagnostic Statistics for the final item set of USIQoL (3 scales, 15 items)

Domain/ Individual Item	Location	Fit residual	Chi Sq.	'p' value	*DIF ['p' values for different patient traits]					
I) PPH (Pain and Physical health)	Value/(with sub-testing)				Age groups	Sex	Symptoms	Site	Treatment Type	Previous Stones
1. Pain	0.34	-1.84	7.75	0.26	0.76	0.49	0.14	0.04	0.55	0.33
2.Pain and physical activity	0.21	-2.32	3.87	0.69	0.13	0.27	0.74	0.50	0.03	0.17
Subtest (1+2)	-0.09	-1.5	4.67	0.59	0.08	0.59	0.09	0.37	0.03	0.08
3. Urgency	-0.4/ 0.25	1.22/-2.0	10.13/6.2	0.12/0.4	0.29	0.75	0.68	0.49	0.37	0.42
4. UTI	0.23/-0.4	-1.40/1.3	3.33/9.2	0.77/0.17	0.05	0.31	0.82	0.94	0.58	0.28
5. Appetite	0.59	-2.06	6.59	0.36	0.46	0.29	0.86	0.19	0.05	0.46
6. Energy	-0.35	-2.5	10.34	0.11	0.10	0.08	0.95	0.03	0.52	0.72
Subtest (5+6)	0.20	-1.35	9.78	0.13	0.26	0.31	0.87	0.54	0.15	0.04
II)PSH (Psycho-Social health)										
7. Sleeping	0.04/0.16	-1.0/-1.1	3.6/8.04	0.73/0.24	0.45	0.05	0.67	0.14	0.04	0.05
8. Depressed	0.19/0.37	-1.0/-1.64	11/4.02	0.09/0.67	*0.007*	*0.004*	0.22	0.22	0.17	0.89
9. Future	-0.17/0.4	0.48/-0.9	3.05/3.5	0.80/0.75	0.40	0.10	0.11	0.27	0.98	0.74
10. Journey	-0.53/0.6	-1.47/0.3	20.6/2.7	0.00/0.84	*0.001*	*0.001*	0.25	0.11	0.30	0.04
11. Help	0.52/-0.6	-0.39/-0.9	4.2/10.8	0.65/0.09	*0.007*	0.05	0.43	0.64	0.02	0.49
12.Medications	0.17	-1.64	10.49	0.11	0.25	0.57	0.95	0.87	0.41	0.06
13. Water	-0.54	0.86	4.81	0.57	0.55	0.03	0.34	0.08	0.09	0.27
Sub-test (12+13)	0.56	0.06	4.30	0.64	0.20	1.0	0.36	0.69	0.10	0.26
III)Work										
14. Time off	0.22	-0.93	5.66	0.23	0.70	0.23	0.81	0.48	0.76	0.99
15. Ability	-0.22	-1.34	3.72	0.45	0.73	0.42	0.73	0.50	0.74	0.95

*DIF (Differential Item Functioning): To establish if the items worked in the same way irrespective of different patient trait