

Development of a disease specific ureteric calculus Patient Reported Outcome Measurement (PROM) instrument.

Maxine GB Tran^{1,2,3}, Michal K Sut^{1,4}, Jane Collie¹, Joana B Neves^{2,3}, Sami Al-Hayek¹, James N Armitage¹, Dominique-Laurent Couturier⁵, Oliver J Wiseman¹

¹Department of Urology, Addenbrooke's Hospital, Cambridge University, Hills Road, Cambridge CB2 0QQ, UK

²Division of Surgery and Interventional Science, University College London, Royal Free Hospital, Pond Street, London NW3 2QG

³Specialist Centre for Kidney Cancer, Royal Free Hospital, Pond Street, London NW3 2QG

⁴ Department of Urology, Peterborough City Hospital, Edith Cavell Campus, Bretton Gate, Bretton, Peterborough, Cambridgeshire, PE3 9GZ

⁵Cancer Research UK – Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE

Running title: Development of a ureteric stone PROM

Running title: Development of a ureteric stone PROM

Maxine Gia Binh Tran * corresponding author

m.tran@ucl.ac.uk tel no: (0044) 0203 002 8787

Michał Sut

michalsut@yahoo.co.uk tel no: (0044) 07704977055

Jane Collie

Jane.collie@addenbrookes.nhs.uk tel no: (0044) 01223 245151

Joana Briosa Neves

j.neves@ucl.ac.uk tel no : (0044) 0203 002 8787

Dominique-Laurent Couturier

Dominique-Laurent.Couturier@cruk.cam.ac.uk tel no: (0044) 01223 769710

Sami Al-Hayek

Sami.hayek@addenbrookes.nhs.uk tel no: (0044) 07572 114180

James Armitage

Jim_armitage@hotmail.com tel no: (0044) 07980 004883

Oliver Wiseman

ojwiseman@gmail.com tel no: (0044) 01223 245151

Abstract

Introduction:

Patient reported outcome measures (PROMs) are powerful instruments to assess the impact of a disease on health from the patient's perspective. We describe the process of designing, testing and validating the Cambridge ureteric stone PROM (CUSP).

Material and Methods:

Patients recently diagnosed with ureteric stones were approached for participation in focus groups, structured interviews, and test-re-test validation studies. Statistical tests included Cronbach's alpha for internal consistency, and Spearman's and Pearson's correlation coefficients for test-re-test validity, permutation tests of equality of means and Spearman's correlation coefficients for discriminant validity.

Results:

Forty-three patients participated in the development of the CUSP. Twenty-two patients were involved in the focus groups and structured interviews and a further 21 participated in the prospective test-re-test study. Expressed comments were grouped into 7 broad health domains: pain, fatigue, sleep disturbance, work and daily activities, anxiety, gastro-intestinal (GI) symptoms and urinary symptoms. Items were selected from established PROM platforms to form the draft (dCUSP) instrument which was then used for test-re-test validation and item reduction. All domains scored highly for Cronbach's alpha (>0.8), with the exception of GI symptoms. Large Spearman's (> 0.76) and Pearson's correlation estimates (>0.83) were obtained for test-re-test validity, suggesting that answers were reliable through the time period tested. The estimates of the Spearman's correlation coefficient between each pair of domains ranged from 0.17-0.78 and the upper bounds of the corresponding 95% confidence intervals were all smaller than 0.95, suggesting that each domain measures something different. The tests of equality of means of scores of the control (n=25) and patient groups were all significant, suggesting that CUSP successfully discriminated patients suffering from ureteric stones for every domain.

Conclusion:

CUSP is a patient derived ureteric stone PROM which can be used to measure ureteric stone disease health outcomes from the patient's point of view.

Introduction:

Outcomes from surgery have traditionally and conventionally been based on surgeon-reported outcomes such as estimated blood loss, infection rates and length of hospital stay. Recently, increased awareness about the need to assess healthcare from output to outcome in a multidimensional way, and the need for better qualitative assessment has led to a growing interest in assessing outcomes from a patient's perspective. Patient reported outcome measures (PROMs) are an effective way to do this(1). PROMs are completed by patients before and after treatment, and can be a powerful measure of health and treatment success. Broad clinical outcomes such as mortality rates inform very little about the overall quality of care(2). PROMs are distinct from patient reported experience measures (PREMs) which focus instead on experience of care such as being treated with dignity, or clinic waiting times(1).

The private hospital group Bupa, now owned by Spire Healthcare, has been routinely collecting PROMs since the 1990s. The National Health Service (NHS) PROMs programme initiated in 2009 by the Department of Health (now NHS England) has required the routine measurement of PROMs for all NHS patients in England before and after receiving surgery for four specific treatments (hernia, hip and knee replacement, and varicose veins)(3), and there is considerable interest for extending measurements to encompass all elective surgery(2). There are two general types of PROMs, those that are generic (e.g. EQ-5D and SF-36) and measure quality of life indicators(4, 5), and those that are disease specific (e.g. International prostate symptom score (I-PSS) (6) and urethral stricture PROM(7)) which measure specific symptom indicators to assess impact of the disease in question.

Although PROMs have been used in Urology (7, 8), such as the Ureteral stent symptom questionnaire (9), and the Wisconsin Stone Quality of Life Questionnaire (WISQOL) has recently been validated in kidney stone formers (10-12), a systematic review of the literature did not identify a specific PROM for ureteric calculi; one of the commonest urological emergencies. Here, we describe the process of designing, testing and validating a disease specific ureteric stone PROM, the Cambridge ureteric stone PROM (CUSP).

Material and methods:

Stage 1.

Patients recently diagnosed or treated for ureteric stones were identified at the stone multidisciplinary team (MDT) meeting and stone clinic in a tertiary referral stone unit, Cambridge University Hospitals NHS Trust, and approached for participation in the development of CUSP. NHS Research Ethics Committee (REC) approval was granted (REC 14/N1/1111). Patients with kidney stones, ureteric stents, urinary tract infections or pre-existing lower urinary tract conditions were excluded from the study. All patients were given an information sheet explaining the basis of the research, and provided written informed consent. Recruited patients were invited to either attend a focus group or a structured interview session (13, 14). Each focus group was comprised of at least 3 clinicians/research team members (moderator, assistant moderator and transcriber). To facilitate the session, the moderator would use a standard set of open-ended questions that were identified by the researchers such as:

1. What impact has having a stone had on your general health?
2. What symptoms related to the stone did you find bothersome?
3. What impact has having a stone had on your general day to day activity?
4. What impact has having a stone had on your work performance?
5. What impact has having a stone had on your home life?
6. What other issues do you feel are important and that have not been addressed?

These open-ended questions also formed the basis of the one-to-one structured interviews. All focus group and structured interview sessions were audio recorded. Each of the patients' comments were then anonymised and transcribed. Content analysis of the transcribed comments identified recurring patient reported 'themes'. These were then tabulated to indicate frequency of reports and grouped into appropriate broad health domains such as 'urinary symptoms', 'pain' and 'sleep related'. Additional structured interviews were conducted to ascertain whether the themes adequately encompassed the spectrum of experiences and impact for ureteric disease, until no new comments were given ('content saturation').

Stage 2.

The themes identified from Stage 1 then formed the basis of the question/item selection for the first draft of CUSP. Themes were 'weighted' depending on the frequency of the comments; for example, pain was the most frequently occurring concern and was proportioned to have 8 questions, whereas urinary symptoms, gastrointestinal symptoms and fatigue were equally weighted and were attributed 3 questions selected for each. A pragmatic approach was to select questions from previously reported and well validated PROMs (PROMIS healthcare Organisation and FACIT.org).

A representative subgroup of patients ($n = 10$) who participated in the focus groups and structured interviews were then invited to review CUSP, to test comprehension and to comment on whether it accurately and faithfully represented the issues that they had raised. Duplicated or redundant items were removed.

Stage 3.

The revised CUSP instrument was then evaluated for reproducibility with test-retest validation and discriminant validity tests. Cronbach's alpha was used to assess the strength of the relationship between questions included in the PROM. Pearson's and Spearman's correlation coefficients were used to evaluate the level of association between participants scores measured at different time points. We also assessed whether the different domains (pain, fatigue, etc.) were different from each other, to ensure that each domain measured something different (15). Therefore, we estimated the Spearman's correlation coefficients between the scores of the patient groups for each pair of domains and defined the corresponding 95% percentile confidence by means of a non-parametric bootstrap with $R=10'000$ replicates (16).

Finally, we tested the ability of the CUSP to discriminate controls from patients with by comparing the mean CUSP results of the 21 patients with ureteric stones to 25 controls (without ureteric stone disease) for each domain of CUSP. Due to the relatively small sample size per condition, we used one-

sided non parametric permutation tests with R= 10'000 permutations to test the equality of means between conditions for each domain (16).

Figure 1 is a flow diagram presenting the developmental process of CUSP.

Results:

Stage 1.

Patients with recent acute presentation with ureteric stones participated in 3 focus groups (n=11) and structured interviews (n=11). There were 4 female and 18 male patients, with an age range 34-78 years (median 54.5 years). Fifteen patients were married or had partners, 4 were divorced and 3 were single. Twenty patients were Caucasian, 1 was of mixed race and 1 was of Asian ethnicity. Six patients were educated to bachelor's degree or higher, 9 patients were educated to Advanced ('A') levels and 7 to General Certificate of Secondary Education (GCSE) standards. Ureteric stone size ranged from 3-11mm (median 5mm). Eleven of these patients had Shock Wave Lithotripsy (SWL), 6 had medical expulsive therapy, and 5 had ureteroscopy (URS) and laser fragmentation for their ureteric stones. Patients attended the focus groups or structured interviews between 2-4 weeks after diagnosis, either whilst awaiting or following treatment. In all, 218 patient reported impact comments were transcribed, and these were tabulated to identify the occurrence of recurring reports and comments (Figure 2). These were then grouped into 7 broad health domains: Pain, Fatigue, Work and Daily Activities, Sleep disturbance, Anxiety, Gastrointestinal (GI) symptoms and Urinary Symptoms.

Stage 2.

Items were selected for the 7 broad health domains identified in Stage 1. Pain was the most frequently reported concern and was represented by 8 items, Fatigue was the next most common and was proportioned to have 5 items, followed by 4 items allocated to Sleep disturbance, and 3 items to each of Work and Daily activities, Anxiety, GI symptoms and Urinary symptoms. A pragmatic approach was used to select and adopt previously well validated questions from established PROM instruments (PROMIS healthcare Organisation and FACIT.org). These 29 items consisted of statements followed by 5-Level Likert scale responses. The total score from this draft version of CUSP (dCUSP) ranged from 29-145 in total. Patients took between 2-3 minutes to complete dCUSP.

Stage 3.

The dCUSP instrument was then used in a pilot prospective test on 21 patients with radiologically confirmed ureteric calculi. There were 6 female and 15 male patients, with an age range 20-67 years (median 50 years). Ureteric stone size ranged from 3-11mm (median 5mm). Seven patients were having SWL treatment, 3 were awaiting or were post treatment with URS and 11 were being treated conservatively. The patients completed dCUSP on two occasions, separated by at least an hour. Cronbach's alpha estimates and dimension reduction techniques using principal component analyses suggested that, for all domains *except* GI symptoms, the questions related to each domain measured the same (underlying) construct or variable (Table 1). The low Cronbach's alpha estimate obtained for

GI symptoms was mainly due to the use of a ‘reverse scale’ in one of the 3 questions related to the domain. Removal or reversing the scale of the question resulted in a Cronbach’s alpha score of 0.52, likely due to some of the patients correctly reading the scale while others did not. In view of the ambiguity of the construct and the relatively low importance of the domain compared to others, it was removed from the final CUSP instrument. Of note, Cronbach’s alpha score in all the domains (apart from GI symptoms) were larger than 0.7, which is often used as the lowest acceptable limit.

For test-re-test validity, Spearman’s and Pearson’s correlation estimates were used. Large values were obtained when analysing patients average scores at time points 1 and 2 for each domain (Figure 3), suggesting that the answers were reliable through the time period tested.

We assessed discriminant validity by comparing the averages of the scores of participants to the control and patient groups for each domain of the CUSP. Figure 4 shows the box plots of the scores of participants per study allocation group and domain, as well as the average of the scores per condition (pink dots). The mean scores of patients were systematically larger than the ones of controls. The analysis of the medians (thick black lines) led to the same conclusion. For the domain ‘Anxiety’, the median score of patients was only slightly higher than the median score of the control group but as its distribution was more skewed, the means per group were different. Also, for all domains, the spread of the scores of the patient group seemed larger than the one of the control group. Table 2 reports the mean difference between patient and control group scores per domain (mean of the patient group minus the mean of the control group), the 95% confidence intervals (CI) for the difference in means as well as the p-values of the one-sided permutation tests of equality of means.

All tests were statistically significant and no confidence interval contained 0. Therefore, we could conclude that the proposed CUSP can successfully discriminate patients with ureteric stones from controls for every domain.

Finally, we assessed if each construct measured something different to the others by estimating Spearman’s correlation coefficient between pairs of domains (Table 3), and by defining the corresponding 95% confidence interval (Table 4). Given that the estimated Spearman’s correlation coefficient estimates ranged between 0.17-0.78 and that the upper bounds of the corresponding 95% CI were all below 0.95, we concluded that each domain measured something different.

Discussion:

Donald Berwick eloquently quoted in 1997 that ‘The ultimate measure by which to judge the quality of a medical effort is whether it helps patients (and their families) as they see it. Anything done in health care that does not help a patient or family is, by definition, waste, whether or not the professions and their associations traditionally hallow it’(17). Patient reported outcome measures are valuable tools from which clinicians and healthcare providers can gain important insights into the effects of medical and surgical interventions specifically from the patient’s point of view. It is now widely recognised that classical clinical outcomes alone provide little information about what is relevant to patients. For example, a surgeon may declare an operation to be a success, but this means very little if the patient reports little benefit following surgery (2, 18). Case in point, data from the NHS PROMs programme showed that by using a generic PROM such as the EQ5D almost half (49.8%)

of patients reported improvement of their overall health following a groin hernia repair; while 88.4% of patients reported improvement following hip replacement. Use of a disease specific PROM (Oxford Hip Score) pre and post treatment, showed an even more impressive 96.5% improvement rate in these patients across the nation (19)(<http://digital.nhs.uk/catalogue/PUB30036>). Therefore, while use of a generic PROM is useful to measure universal health issues, disease specific PROMs such as the Oxford Health Score enables measurements of the effect of treatment on specific aspects of the disease in question.

Although the urology community has recognised the value of PROMs and embraced their use as a valuable means to benchmark treatment outcomes, there remains a paucity of urological disease specific PROMs. The most widely used example is the I-PSS which is adopted in standard practice to inform clinicians of the patient impact of prostatic symptoms before, during and following treatment (medical or surgical) (6). A PROM for urethral stricture surgery has also been successfully validated and is used internationally in several centres (20-22). More recently, exciting progress in PROMs has been made in the area of kidney stones. The WISQOL is a PROM for patients with a history of kidney stones. Its external validation in a multi-centre study in North America showed that although the results correlated with the generic SF-36v2 (Pearson correlation 0.56, p<0.001), the WISQOL was a superior instrument in discriminating between patients with and without current kidney stones (10, 11). However, WISQOL has yet to be assessed for test-re-test validity and clinical use (10). Joshi and colleagues have also reported on the development of a ureteral stent symptom questionnaire, to evaluate the impact of different types of ureteric stents (9).

Ureteric calculi are one of the commonest urological emergencies. Various treatment modalities such as SWL, ureteroscopic fragmentation, or medical expulsive therapy are available. Although there are broad clinical guidelines, treatment allocation is variable and depends on clinical judgement, access and patient factors. Current evidence for each of these modalities is largely based on clinical outcome measures. Our objective was to produce a disease specific ureteric stone PROM that could aid in treatment outcome assessment. In this article, we have described the processes involved from initially identifying the need for a new instrument, item generation, item consolidation and reduction and test-re-test validation. We have highlighted the importance of embedding patient involvement throughout the design and refinement stages, enabling the development of a truly patient derived PROM.

There are three main limitations to our study. First, our patient cohort was relatively small, and a distinction was not made between first time stone formers and recurrent stone formers, and so may not be representative of the general population suffering from ureteric stone disease. CUSP (Figure 5) is currently being used in a national multi-centre study, the results of which will inform on demographic, previous history of stone disease, and geographical variances. Second, we selected items from previously validated PROMs available on public platforms, rather than designing our own questions. This was a pragmatic decision, based on the need to work within the constraints of the resources available. Creating a questionnaire from scratch would necessitate psychometric testing and validation, which was beyond the scope of our study. Third, to complete validation of CUSP, it is necessary to demonstrate that the instrument is sensitive and can measure *change* as part of clinical validation studies. This is currently in progress as part of the UK national Therapeutic Interventions for Stones of the Ureter (TISU) study, and we eagerly await the results (expected in 2019).

Having guaranteed patient involvement at every stage of the development of CUSP (Figure 1), a major strength of our tool (23), we are confident regarding its fidelity and utility. CUSP could provide clinicians with important insights into the impact of ureteric stone disease on patients. It will enable for the first time quantitative comparison of qualitative outcomes for management options for ureteric stones. Further validation currently ongoing will confirm if CUSP is an instrument that accomplishes its original aim.

Conclusion:

We describe the detailed process of designing a PROM that is truly patient derived and encourage other clinicians and researchers to look beyond standard clinical indices of success and focus on ways to measure outcome from the patient's point of view.

Acknowledgements:

We would like to thank Ms Angela Cottrell, Mr Andrew Dickinson and Professor Sam McClinton; and all the patients who participated in this study.

Disclosures:

O J Wiseman is in receipt of grants/research support from Porges Coloplast; is in receipt of honoraria or consultation fees from Boston Scientific, EMS and Porges Coloplast; participates in company sponsored speakers' bureau from Boston Scientific, EMS, Porges Coloplast and Olympus; and is a stock shareholder of Uroscreen Ltd.

References

1. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346:f167.
2. Timmins N. NHS goes to the PROMS. *BMJ*. 2008;336(7659):1464-5.
3. Devlin NJ, Parkin D, Browne J. Patient-reported outcome measures in the NHS: new methods for analysing and reporting EQ-5D data. *Health Econ*. 2010;19(8):886-905.
4. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Br J Rheumatol*. 1997;36(5):551-9.
5. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ*. 1992;305(6846):160-4.
6. Bosch JL, Hop WC, Kirkels WJ, Schroder FH. The International Prostate Symptom Score in a community-based sample of men between 55 and 74 years of age: prevalence and correlation of symptoms with age, prostate volume, flow rate and residual urine volume. *Br J Urol*. 1995;75(5):622-30.
7. Jackson MJ, Sciberras J, Mangera A, Brett A, Watkin N, N'Dow J M, et al. Defining a patient-reported outcome measure for urethral stricture surgery. *Eur Urol*. 2011;60(1):60-8.
8. Tran MG, Yip J, Uveili K, Biers SM, Thiruchelvam N. Patient reported outcome measures in male incontinence surgery. *Ann R Coll Surg Engl*. 2014;96(7):521-5.
9. 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. *The Journal of urology*. 2003;169(3):1060-4.
10. Penniston KL, Nakada SY. Development of an instrument to assess the health related quality of life of kidney stone formers. *J Urol*. 2013;189(3):921-30.
11. Penniston KL, Antonelli JA, Viprakasit DP, Averch TD, Sivalingam S, Sur RL, et al. Validation and Reliability of the Wisconsin Stone Quality of Life Questionnaire. *J Urol*. 2017;197(5):1280-8.
12. Penniston KL, Nakada SY. Use of the WISQOL Questionnaire. *J Endourol*. 2017;31(4):420.
13. Rothrock NE, Kaiser KA, Cella D. Developing a valid patient-reported outcome measure. *Clin Pharmacol Ther*. 2011;90(5):737-42.
14. Wiering B, de Boer D, Delnoij D. Patient involvement in the development of patient-reported outcome measures: a scoping review. *Health Expect*. 2017;20(1):11-23.
15. Hair JF BW, Babin BJ, Anderson RE. Multivariate Data Analysis: A Global Perspective. 7th ed: Prentice Hall, Upper Saddle River; 2009, page 680.
16. Davison AC HD. Bootstrap Methods and Their Application . Cambridge University Press, Cambridge; 1997, section 4.3-5.3.
17. Berwick DM. Medical associations: guilds or leaders? *Bmj*. 1997;314(7094):1564-5.
18. Meredith P, Emberton M, Devlin HB. What value is the patient's experience of surgery to surgeons?: the merits and demerits of patient satisfaction surveys. *Ann R Coll Surg Engl*. 1993;75(3 Suppl):72-3.
19. Lim CR, Harris K, Dawson J, Beard DJ, Fitzpatrick R, Price AJ. Floor and ceiling effects in the OHS: an analysis of the NHS PROMs data set. *BMJ Open*. 2015;5(7):e007765.
20. Puche-Sanz I, Martin-Way D, Flores-Martin J, Exposito-Ruiz M, Vicente-Prados J, Nogueras-Ocana M, et al. Psychometric validation of the Spanish version of the USS-PROM questionnaire for patients who undergo anterior urethral surgery. *Actas Urol Esp*. 2016;40(5):322-7.
21. Jackson MJ, Chaudhury I, Mangera A, Brett A, Watkin N, Chapple CR, et al. A prospective patient-centred evaluation of urethroplasty for anterior urethral stricture using a validated patient-reported outcome measure. *Eur Urol*. 2013;64(5):777-82.
22. Lucas ET, Koff WJ, Rosito TE, Berger M, Bortolini T, Neto BS. Assessment of satisfaction and Quality of Life using self -reported questionnaires after urethroplasty: a prospective analysis. *Int Braz J Urol*. 2017;43(2):304-10.

