Development and preliminary validation of a diagnostic prediction model to optimise outpatient management of patients with urolithiasis using urinary stones and intervention quality of life (USIQoL) measure

Ruth C. Jarvis[1,3] · Philip Pallmann[2] · Collett Clements[1] · Hrishikesh Joshi[1,2]

Abstract

Purpose Patients with urinary calculi undergo resource-intensive follow-up. Application of a PROM, Urinary Stones and Intervention Quality of Life (USIQoL), can potentially optimise current practices if it matches the outcomes of traditional follow-up. Our objective was to develop, and conduct, a preliminary validation of the USIQoL based prediction model to aid triage.

Methods We performed a two phase prospective cohort study. The 1st phase included development of the USIQoL-based decision model using multicentre data. The 2nd phase involved prospective single-blind external validation for the outpatient application. The aim was to evaluate correlations between the USIQoL scores and key predictors; clinical outcomes and global health ratings (EuroQoL EQ-5D). We used statistical analysis to validate USIQoL cut-off scores to aid triage and the decision to intervene.

Results Of 503 patients invited, 91% (n = 455, Development [305] and Validation [150]; M = 308, F = 147) participated. The relationship between USIQoL domain scores and clinical outcomes was consistently significant (estimated odds: PPH 1.24, p < 0.001, 95% CI 1.13–1.36; PSH 1.22, p < 0.001, 95% CI 1.12–1.33). The ROC values for the model were ≥ 0.75. The optimum domain cut-off scores were derived with rising scores implying increased need to intervene. The model demonstrated satisfactory sensitivity (0.81–0.89) and specificity (0.36–0.47).

Conclusions The study demonstrates satisfactory correlation between the USIQoL and clinical outcomes making this model a valid aid for triage and optimising outpatient management with the cut-off scores able to identify high risk patients who need active treatment.

Keywords Urolithiasis · Prediction · Follow-up · MCID · PROM · USIQoL

Introduction

Urolithiasis has a global prevalence of around 10 percent of the population [1]. The burden of the disease is increasing, along with its associated economic impact [2]. The healthcare cost of treating urolithiasis is comparable to the combined cost of bladder and prostate cancer in the UK [3]. Most patients with urolithiasis undergo long-term follow-up involving regular clinic reviews and imaging to prevent or identify possible complications early. This is resource-intensive, involves exposure to ionising radiation and is not without diagnostic limitations. Furthermore, there are wide variations in practices. Deciding the optimal frequency and duration of follow-up is a longstanding problem with little

✉ Hrishikesh Joshi
  hrishijoshi@hotmail.com

Ruth C. Jarvis
  ruth.jarvis@wales.nhs.uk

1 Urology, Cardiff and Vale University Health Board, Cardiff, UK

2 Centre for Trials Research, Cardiff University, Cardiff, UK

3 Department of Urology, University Hospital of Wales, Heath Park, Cardiff, Wales CF14 4XW, UK

evidence base and alternatives. The National Institute for Health and Care Excellence (NICE) indicated that currently no recommendations can be made regarding follow-up and that more research is needed [4]. The European Association of Urology urolithiasis guidelines panel have published guidance on the follow-up strategy for patients undergoing definitive treatment, based on different stone, patient and imaging factors. They have commented that the evidence base remains limited due to the high heterogeneity and the lack of any comparative data [5].

Recently, there has been a shift to patient-centred healthcare, and as a part of that, Patient Reported Outcome Measures (PROMs) have been developed which seek to ascertain patients' views of their symptoms, functional status, and well-being [6]. Recent studies support the use of PROMs in clinical practice for improved shared decision-making and patient self-management [7]. They have been found to be used to "identify triggers for surgery and potentially reduce the burden on services by limiting unnecessary or ineffectual procedures" [8]. When used on a longitudinal basis, PROMs can track the progression and severity of disease and be incorporated as an adjunct to make changes to treatment and follow-up [9]. The American Urological Association (AUA) guidelines state that treatment decisions about urinary calculi should incorporate patient preferences that are influenced by the Health-Related Quality of Life (HRQoL) impact rather than the limited clinical and radiological outcomes [10].

Urinary Stones and Intervention Quality of Life (USIQoL) has been specifically developed as a core PROM for patients with upper urinary tract stones. It is scored in 3 sections covering the domains of Pain and Physical Health (PPH, 6 items), Psycho-Social Health (PSH, 7 items) and Work performance (W, 2 items). It has a completion time of 3–4 min and is well suited for longitudinal application [11]. Its completion involves patients rating the amount of bother attributed on a 4-point scale (1 = not at all, 2 = a little, 3 = quite a bit or 4 = a lot). The scale scores are generated by a simple summation of scores for each item in the domain (score range: PPH 6–24, PSH 7–26 points, W0-8) with higher scores indicating greater patient bother. The results from its validation study demonstrated that USIQoL is reliable (r ≥ 0.8), internally consistent (α ≥ 0.7) with good construct validity (good hypothesised correlations, r > 0.3) and sensitivity to change (p < 0.01). All scales demonstrated unidimensionality with good item fit and person separation indices.

When a patient with urolithiasis attends a clinic, the important question is whether there is a need for additional tests, and possible intervention. This is even more important in the context of patients with small and recurrent stones attending regular follow-up. In this setting, it would be important to know if the adoption of the USIQoL

as a monitoring tool into routine practice can assist clinical decision-making, if the results correlate well with those of traditional methods of outpatient review, thus serving as an alternative way to manage the patients.

Our study had the following objectives:

1) To develop the first in endourology, PROM (USIQoL) based prediction model to identify patients at risk of needing intervention and conduct a preliminary validation of the model for outpatient use.
2) To develop USIQoL cut-off scores, for the physical and psycho-social domains, that can reliably differentiate between patients needing traditional evaluation and possible intervention(s) against the low risk stable group suitable for PROM only follow-up.
3) To establish the Minimal Clinically Important Difference (MCID) for the USIQoL defined as "the minimal change in the score considered to be relevant by patients and physicians" [12].

Materials and methods

The study received ethical approval from the Southeast Wales research ethics committee (17/WA/0195) and the local governance committee for service evaluation. It was performed in accordance with the Declaration of Helsinki. Established methodology for development of a prediction model was followed to derive clinically significant changes in the USIQoL scores and cut-off values [13].

The study involved the administration of two PROMs, USIQoL (disease specific) and EuroQoL EQ -5D-3L (generic) to patients during their outpatient attendance with the results analysed using summary index scores [14]. We present the results based on the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement [15].

The study was conducted in 2 phases.

Phase I—development of a USIQoL-based prediction model

This involved the analysis of the PROM scores and clinical outcomes, from an existing dataset collected in a prospective multicentre (4 secondary care urology departments) cohort study during the final phase of the validation of the USIQoL questionnaire that is already published [11]. This phase involved recruitment from March 2018 to June 2019 with the follow-up completed in September 2019. The study included patients with urolithiasis in both the outpatient and inpatient (including emergency) settings who completed the PROMs questionnaires prior to their face to face assessments. The analysis included evaluation of the outcomes stated below

(A & B) and the development of suitable cut-off scores to distinguish between patients requiring intervention (high risk) or not.

Phase II—prospective single-blind preliminary external validation of the prediction model

We tested the decision model for an outpatient application, using a blind single centre prospective study with the data collected over 10 months (December 2021–October 2022). This external validation involved a separate sample (to the development phase) from a later period. Patients with urolithiasis attending urology outpatient clinics were invited to complete both questionnaires just prior to their clinic review. The completed questionnaires were collected by an independent clinic staff member not involved with the subsequent outpatient review. The patients underwent complete assessment by the clinician including a review of their most recent imaging. A final decision regarding management was formulated, with shared care decision-making, and documented in the case notes. The clinicians were blinded to the questionnaire scores. The validity of the Phase I model, including the proposed USIQoL cut-off scores, was subsequently evaluated. The eligibility criteria, predictors and the outcome data assessed were similar to the development phase.

The predictors, for which the data was collected as well as inclusion and the exclusion criteria, are shown in Table 1. The predictors used are part of the standard protocols followed in the daily practices across all the urological departments.

Our outcomes involved looking at the relationship between the USIQoL domain scores (PPH and PSH) and the outcomes from:

1) Traditional face to face assessments using two key clinical parameters (A1 and A2 below) and
2) EQ-5D-3L scores when administered simultaneously.

The clinical predictors were:

A1) Decision regarding clinical management: This was based on the face to face patient assessments involving history taking and clinical examination as well as imaging studies. It was categorised into outcome A (active interventional treatment, which included shockwave lithotripsy, ureteroscopy, percutaneous nephrolithotomy or medical therapy with curative intent, such as dissolution treatment), and outcome B (no intervention, suitable for simple follow-up).
A2) We separately looked at the additional predictor of presence of ongoing symptoms, at the face to face assessment, attributable to the stones.
B) The results of the EQ-5D-3L and EQ-5D-3L Visual Analogue Scale (VAS) scores served as an external independent variable to assess global health.

For subjective measures in general, including the studies and models involving application of the PROMs, the Food and Drug Administration (FDA) recommends different types of anchors or predictors, as external criteria approximating truth. These also help to generate relevant thresholds for meaningful within-patient change. These serve as tools for internal validation and are useful for the cross checking of the outcomes. Those used were (1) Established clinical outcomes (outcomes A or B in our case), (2) Global impression of change (development of stone-related symptoms, 'yes or no', a binary variable), and (3) Current-state global impression of severity (EQ-5D PROM index scores in this

Table 1 Inclusion/exclusion criteria and predictor data

Inclusion	≥ 16 years old Renal and/or ureteric calculi requiring review in the secondary care in the inpatient or outpatient setting	
Exclusion	Unable to self-administer the questionnaire Lack of mental capacity Pregnancy	
Predictor data	Demographics	Age, sex
	Clinical information	Stone symptoms, relevant prior history (previous and current treatments), physical assessment, imaging*
	Clinical outcomes	Outcome A–active interventional treatment** Outcome B–no active intervention***
	Radiological information (to decide clinical outcomes A and B)	Stone site, stone size, stone number, change/s when compared with recent/previous imaging
	USIQoL scores	PPH and PSH domains
	EQ-5D-3L	Utility and thermometer(global rating) score

*Protocol based: X-rays for radiopaque stones, Ultrasound for lucent stones, Non-Contrast CT at baseline for all and subsequently (if deemed necessary on clinical grounds); **Active interventional treatment = shockwave lithotripsy, ureteroscopy, percutaneous nephrolithotomy or medical therapy with curative intent; ***No active intervention = follow-up with or without subsequent discharge (first time, low risk)

study, a rank variable with higher scores indicating better health state) [16]. For the purposes of this model, USIQoL scores from the 2 major domain scales (PPH and PSH—13 questions) were considered. The work domain (paid employment) was not included as it was not applicable to all of the patients uniformly.

Sample size considerations

We followed the rule of thumb sample size recommendation; a) for psychometric analyses of summated scales (PROMS), based on the guidance for their traditional validity assessments [minimum 10 subjects per item of the total scale items (13 items for 2 domains in the USIQoL)], b) to ensure at least 10 events for each predictor parameter (widely referred to as events per variable EPV) being considered for inclusion in the prediction model [17].

For the development phase, we had data from a sample representative of the secondary care urolithiasis patient population, of 305 patients (with minimum requirement of 130 patients) from the 4 centres. We decided to use all available data to maximise the power and generalisability and cover a higher outcome proportion. For the preliminary validation, we included the data from 150 patients (> minimum 130 patients). As the PROMs were completed by the patients in the hospital prior to face to face interaction, we did not encounter problems with missing PROM data. The data for all other predictors was available as part of the normal clinical practice. The number of patients who did not enter the study remained small.

Statistical analysis

The analysis was undertaken using SPSS (version 29) software. Spearman's Rank Correlation Coefficients were used to calculate the correlations between PPH and PSH and EQ-5D scores. We used Binomial Logistic Regression to assess the strength of the relationship between the two domain scores (PPH and PSH) with the clinical outcomes (A or B), stone-related symptoms and radiological parameters (stone site, size). We used Receiver Operating Characteristic (ROC) analysis to assess discrimination between outcomes A and B by looking at the area under the curve (AUC). An AUC of 0.7–0.8 is considered acceptable [18]. We also calculated the sensitivities and specificities for different cut-off scores.

We then selected the optimal cut-off score by first evaluating Youden's Index [19]. It provides the best trade-off between sensitivity and specificity. In our study, if the sensitivity with the highest Youden's index was below 0.70, we deemed it unacceptable and chose the nearest threshold with a sensitivity above 0.70. We calculated the cut-off scores for the two domains independently as well

as when applied in combination (being clinically most relevant). We determined the MCID using a combination of anchor-based and distribution-based statistical methods [20].

Results

Of the 503 patients who were invited to complete the study during the 2 phases (Phase I = 345 and II = 158), 455 (90.3%) patients participated (Phase I = 305, Phase II = 150). We have presented TRIPOD 22 point checklist as Appendix I. The results are shown in Table 2. There was a male preponderance in both phases with a similar mean age and age range. In phase I, a higher number of patients required intervention. The differences in distribution of stone site, and the proportion of clinical outcomes between the two phases was attributed to the fact that the main focus in phase II was outpatients, with a relatively smaller number of patients suffering from acute stone episode.

Table 3 demonstrates the distribution of the domain scores, across phases, for both clinical outcome predictor groups (A and B). The mean and median values for the PPH and PSH domains in both data sets are higher in the intervention, compared with the no-intervention group, as expected. There was significant difference between the scores, for both domains, between intervention and the non-intervention group ($p < 0.05$, Table 3). The phase II data demonstrated that renal stones under and above 11 mm, were 88% and 12% respectively (total renal stone size range 3 mm to partial staghorn). 42% ureteric stones were under 5 mm, 42% between 5 and 10 mm and 16% were above 10 mm (range 3 mm–15 mm). Both domain scores for patients with ureteric stones were higher when compared with renal stones. The mean PPH score for ureteric stones was 13.8 and 10.6 for renal stones, and the mean PSH for ureteric stones was 14.9 when compared with 12.5 for renal stones.

The relationship between the PPH and PSH domain scores with the important clinical predictor anchors (outcomes A or B and presence of symptoms) was statistically significant (Table 3). From this, it was clear, that the odds of the patient expressing symptoms of stones, and needing full clinical evaluation and subsequent active intervention, increased with the increasing USIQoL scores. The results confirmed good correlation and one-dimensionality between the two domains.

Figures 1 and 2 show the ROC curves, and the derived AUC values, for PPH and PSH for phase II. All of the potential cut-off values, including those for Phase I PPH and PSH, along with their accompanying sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR) and Youden's Index, are shown in Table 2. The AUC values

Table 2 Descriptive statistics

		Development data (Phase I) N = 305	Validation data (Phase II) N = 150
Demographic data	Age [range], yrs	53.8 [18–90]	57.51 [19–91]
	Gender	M = 215 F = 90	M = 93 F = 57
Radiological data (n = patient no) and size range (smallest, when visible – largest)	Renal	192 (3 mm–Staghorn)	110 (2 mm–partial Staghorn)
	Ureteric	113 (3 mm–14 mm)	12 (3 mm–16 mm)
	Not visible clearly	0	28
Co-morbidities data [American Society for Anaesthesiology (ASA)]	ASA I	28%	35%
	ASA II	40%	37%
	ASA > III	32%	28%
Prior history of urolithiasis		46%	65%
Clinical outcomes	ESWL	56%	54%
	Ureteroscopy	29%	25%
	PCNL	6%	18%
	Medical management only	9%	3%
	No intervention (follow-up)	89	111

M male, *F* female

Table 3 Statistical analysis of domain scores

|	Development (Phase I)		Validation (Phase II)	
	PPH score	PSH score	PPH score	PSH score
No intervention				
Mean and SD	9.13 ± 3.36	10.81 ± 2.80	9.43 ± 3.82	11.41 ± 4.14
Median	8	9	8	10
Active interventional treatment				
Mean and SD	11.14 ± 4.24	12.94 ± 4.03	13.74 ± 4.83	15.74 ± 4.74
Median	10	12	15	15
Logistic regression (estimated odds)				
Symptoms	1.203 95% CI 1.117–1.296 p < 0.001	1.101 95% CI 1.007–1.203 p = 0.034	1.598 95% CI 1.390–1.837 p < 0.001	1.599 95% CI 1.382–1.849 p < 0.001
Clinical outcome	1.148 95% CI 1.063–1.240 p < 0.001	1.179 95% CI 1.020–1.363 p = 0.025	1.240 95% CI 1.133–1.357 p < 0.001	1.220 95% CI 1.120–1.329 p < 0.001
Internal consistency	PPH scale	PSH scale	PPH scale	PSH scale
Cronbach's alpha	0.82	0.75	0.87	0.84
Correlation with EQ-5D (Spearman's rank coefficients)				
EQ-5D-3L converted utility score	r_s −0.471 95% CI (−0.587)–(−0.337) p < 0.001	r_s −0.494 95% CI (−0.683)–(−0.242) p < 0.001	r_s −0.613 95% CI (−0.713)–(−0.488) p < 0.001	r_s −0.573 95% CI (−0.682)–(−0.440) p < 0.001
EQ-5D-3L VAS	r_s −0.407 95% CI (−0.538)–(−0.258) p < 0.001	r_s −0.186 95% CI (−0.451)–(0.109) p = 0.201	r_s −0.663 95% CI (−0.752)–(−0.550) p < 0.001	r_s −0.603 95% CI (−0.705)–(−0.476) p < 0.01

were 0.75 (PPH) and 0.76 (PSH) in phase II, which demonstrated the satisfactory ability of the model to differentiate between the two clinical outcomes.

The ROC analysis results, together with the sensitivity, specificity, Youden's Index, Positive Likelihood Ratio and Negative Likelihood Ratio for a range of cut-off scores, were

ROC for Phase II PPH
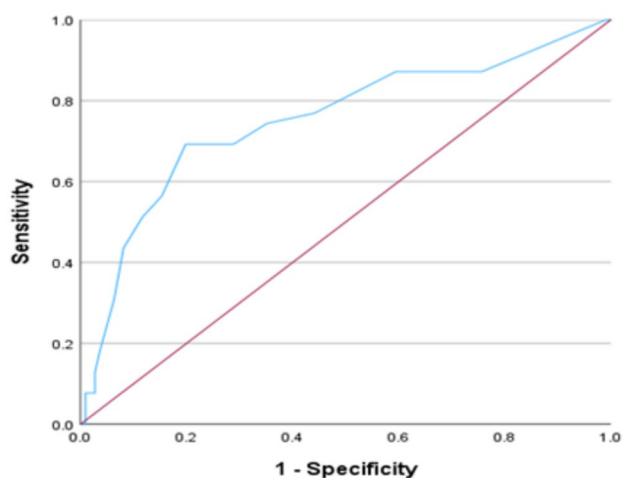
Area Under the Curve = 0.752 (95% CI 0.654–0.849)

Fig. 1 ROC Curve for Phase II PPH

ROC for Phase II PSH
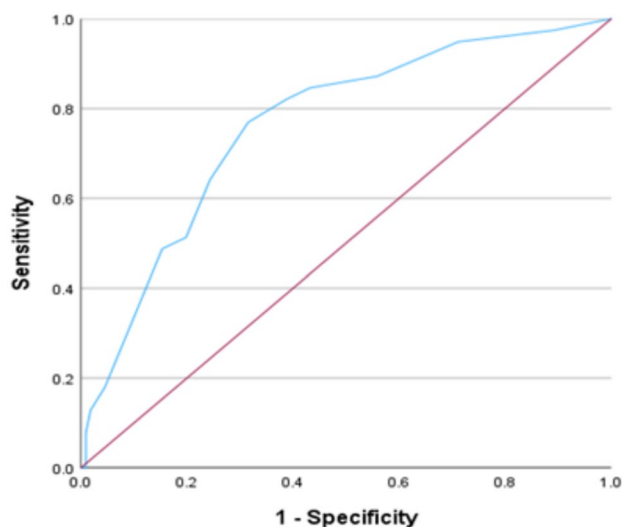
Area Under the Curve = 0.763 (95% CI 0.677–0.849)

Fig. 2 ROC Curve for Phase II PSH

used to determine the most appropriate cut-offs for PPH and PSH. We initially assessed Youden's index, taking note of the domain scores used alone and using a combined rule for both domains together. The cut-off values of 11 (PPH) and 13 (PSH), with the highest Youden index, had sensitivities below 0.7; this was deemed to be clinically unacceptable, and a more appropriate cut-off was chosen (sensitivity minimum of 0.8 and specificity around 0.5). The chosen cut-off scores were the combined rule of PPH 9 and PSH 10

(sensitivity 0.815, specificity 0.468, NLR 1.53, PLR 0.40, Youden's Index 0.283). The combined rule means that the cut-off threshold is reached if either the PPH or PSH domain scores or both are above the chosen cut-off. The full range of cut-off values with the associated sensitivity and specificity can be seen in Table 4.

The application of the cut-off scores during phase II demonstrated good and consistent model fit. There was a very good correlation between the EQ-5D utility scores and USIQoL domains, which provided evidence for an aspect of internal validity of their use in outpatient settings (Table 3). In the outpatient setting (Phase II), of the 150 patients, 12 were found to have ureteric stones, with the clinical decision to intervene in 6 of those. The USIQoL domains scores for all 6 of those requiring intervention were above the cut-offs with 100% (PPH) and 83% (PSH) sensitivity. In this cohort, the sensitivities for the PPH and PSH domain for the renal stones, above cut off were 72% and 88% respectively in keeping with the multifactorial decision making for renal stones. The distribution of the domain scores and the respective clinical outcomes suggested the emergence of two groups in the patients with scores above the cut-offs. The results of the validation sample showed that 65% patients with the PPH domain score, and 55% with the PSH domain scores above 14 required intervention. The similar rates of intervention for the PPH and PSH domains with scores 10–14 were, 26% (PPH) and 33% (PSH) respectively. Patients with domain scores ranging between 10–14 points, are served better with full clinical assessment, including up-to-date imaging, to make a fully informed management plan. However, once the domain scores were above 14, the need for intervention was much higher, thus defining it as the high-risk group of patients.

The analysis, to define an MCID that helps understand the magnitude of the QoL impact and change, using different methods, revealed a 3–4 point difference for each of the domain scores (Table 5). It appeared higher for the stable outpatient population with a lower prevalence of interventions. The MCID appeared to be similar across methods of calculation.

Discussion

This two-phase study on the development and preliminary validation of a PROM based prediction model demonstrates the clinical suitability of adopting USIQoL to aid management of urolithiasis in outpatients. It provides an alternative approach to patient-centric evaluation that is useful for daily practices. The results have demonstrated good correlation between the USIQoL scores and the outcomes of traditional assessments. The proposed cut-off scores indicate its ability

Table 4 Phase I and II ROC curve cut-offs and for PPH and PSH

	Cut-off (≥)	Sensitivity	Specificity	PLR	NLR	Youden's Index
Development (phase I)						
PPH	7	0.900	0.247	1.19	0.41	0.147
	8	0.768	0.442	1.38	0.52	0.210
	9	0.663	0.519	1.38	0.65	0.183
	10	0.542	0.636	1.49	0.72	0.178
	11	0.495	0.727	1.81	0.69	0.222
	12	0.426	0.792	2.05	0.72	0.218
PSH	9	0.873	0.238	1.15	0.53	0.111
	10	0.786	0.524	1.65	0.41	0.310
	11	0.683	0.524	1.43	0.61	0.206
	12	0.579	0.619	1.52	0.68	0.198
Combined rule (PPH &/or PSH)	7/9	0.940	0.182	1.15	0.33	0.122
	8/10	0.836	0.403	1.40	0.41	0.239
	8/11	0.805	0.403	1.35	0.48	0.208
	9/10	0.815	0.468	1.53	0.40	0.283
	9/11	0.770	0.455	1.41	0.51	0.225
	10/12	0.650	0.584	1.56	0.60	0.234
	11/13	0.575	0.701	1.92	0.61	0.276
	12/14	0.480	0.753	1.95	0.69	0.233
Preliminary validation (Phase II)						
PPH	7	0.872	0.243	1.15	0.53	0.115
	8	0.872	0.405	1.47	0.32	0.277
	9	0.769	0.559	1.74	0.41	0.328
	10	0.744	0.649	2.12	0.40	0.393
	11	0.692	0.712	2.40	0.43	0.404
	12	0.692	0.748	2.74	0.41	0.44
PSH	9	0.949	0.288	1.33	0.18	0.237
	10	0.872	0.441	1.56	0.29	0.313
	11	0.846	0.568	1.96	0.27	0.414
	12	0.821	0.613	2.12	0.29	0.434
Combined rule (PPH &/or PSH)	7/9	0.974	0.117	1.10	0.22	0.091
	8/10	0.923	0.279	1.28	0.28	0.202
	8/11	0.923	0.351	1.42	0.22	0.274
	9/10	0.897	0.360	1.40	0.28	0.257
	9/11	0.897	0.460	1.66	0.22	0.357
	10/12	0.846	0.550	1.88	0.28	0.396
	11/13	0.820	0.631	2.22	0.28	0.451
	12/14	0.718	0.694	2.34	0.41	0.412

PLR positive likelihood ratio, *NLR* negative likelihood ratio

to discriminate between key clinical decisions and identify those that are likely to need intervention(s).

There is evidence for the usefulness of the PROMs in clinical practice. At a micro-level, PROMs facilitate the detection of physical or psychological problems and adherence to treatments [21]. PROMs compare favourably with other common clinical measures in terms of reliability [22]. With real-time access to the PROM data, it helps clinicians prioritise topics for discussion along with improved patient-clinician communication [23]. At the meso-level, PROM data can help in comparative effectiveness research and evaluate the impact of interventions [24]. PROM-based follow-up has been utilised in other specialties. The National Health Service (NHS) England national PROMs programme has been using these in the follow-up of patients undergoing hip/ knee replacements since 2009 [25]. Within urology, there are established PROMs to monitor patients with lower urinary tract symptoms. Formal assessment of factors affecting a patient's quality of life should be incorporated into clinical care and utilised to guide treatment decisions,

Table 5 Calculation of the minimal clinically significant difference (MCID) for the two domains using anchor and distribution-based methods

Approach	Method	Development (Phase I)		Validation (Phase II)	
		PPH	PSH	PPH	PSH
Anchor-based					
Between patients score change	Mean score difference between clinical outcomes	2.01	2.13	4.32	4.34
Distribution-based					
SEM	$1.96*SD[\sqrt{(1-r)}]$	3.42	3.86	3.36	3.90
Effect size	$0.5*SD$	2.06	1.97	2.25	2.35

SD standard deviation, *SEM* standard error of Mean

as demonstrated in patients with localised prostate cancer treatment [26].

Lack of application of PROMs in urolithiasis, especially as part of routine clinical practice, is common. In addition, variations in clinical practices and a lack of standard follow-up strategies are long-term problems. Evaluation and follow-up of patients have implications for resources and outpatient waiting times. In the latest Urology Outpatient Transformation guide in the UK, "personalised follow up—patient initiated follow up" and "using remote monitoring" were highlighted as two key components with scope for improved PROM-based follow-up [27]. Following the COVID pandemic, there are pressures for changes to outpatient practices and increased acceptance of alternative methods of follow-up.

With all these considerations, we looked at the development and application of a USIQoL based prediction model applicable in outpatient settings. We hypothesised that its use would provide additional valid and reliable data on the impact of the disease on the patients and the PROM scores would correlate with the clinical outcomes. The results could serve as a guide for triaging higher risk patients who need detailed assessments and likely intervention. We proposed a 2 phase study with the development of a USIQoL-based diagnostic prediction model based on the existing data (phase I), followed by a hypothesis-driven prospective, outpatient-based preliminary validation (phase II) that assessed the suitability of this model and the cut-off scores.

The ability of the PROMs to improve decision-making relies on them accurately capturing the burden of disease or treatment. For the PROMs to be useful in the follow-up of chronic conditions, they must be relevant and actionable. It should state what small changes to the scores mean and when there is a need to act or decide on management plans [21]. Hence, we constructed and reported the results based on the 22 point checklist advocated in the TRIPOD

guidelines. One way to lend meaning and interpretation to the PROM is to dichotomise between values where within-patient changes are considered clinically important ("responders") and those that are not [28]. We followed this approach to devise the principal clinical outcomes, A or B.

The predictors (anchors) used in our study were universally acceptable and clinically relevant. Similarly, we carefully constructed cut-off scores (phase I) with relevant sensitivity analysis. After making certain it satisfied the requirements for the ROC analysis, we undertook prospective preliminary validation study with adequate sample size. The results of the logistic regression analysis confirmed satisfactory relationship between the USIQoL scores and clinical outcomes and helped to draw reliable conclusions.

The principles behind satisfactory sensitivity and specificity values (traditionally used for laboratory tests) are applicable in this setting. The high sensitivities and acceptable specificities for both PPH and PSH confirmed the potential of these cut-off scores. Previous PROM studies have shown that ROC values of around 0.6–0.7 are common and satisfactory for patient evaluation [29]. We thought that a sensitivity of > 0.70 was required given the risks of a cut-off with a lower sensitivity resulting in an inappropriately large number of false negatives (presence of potential stone-related complications but low PROM scores). It was assumed that the risks of false negative results would outweigh any benefits against the risks of false positives (need for additional consultation and/or imaging). Hence, we decided to go for the cut-off scores with higher sensitivity while accepting the relatively modest specificity. The results regarding the appropriate stratification for the cohort of patients with ureteric stones that need intervention, a common pressing clinical question, confirmed the validity of our approach. Ureteric stones with USIQoL scores above the cut offs were at the highest risk of intervention followed by the symptomatic stones with significant QoL impact. The PROM showed good and consistent performance in identifying this high risk group of patients.

The analysis of the cut-off scores identified potential cohorts of patients with an increasing likelihood of detailed investigations (scores above the cut-off) progressing to active intervention (scores ≥ 14).

There is a significant clinical interest in defining the MCID for a given PROM so that the magnitude of the clinical impact, or change, can be understood and standardised. It is well known that MCID is a complex concept with multiple facets and variable results based on the methods used. We used a combination of anchor and distribution-based methods to give the best estimates. Our anchors were easily interpretable, widely used, and well correlated with the USIQoL outcomes. The results were fairly consistent across the methods [30].

There are limitations to the study. It was conducted among the English-speaking population. The development of USIQoL in other languages and in other healthcare settings will need further evaluation. Although, the study involved a satisfactory number of patients based on the sample size requirements, it remains relatively small for the final step of large-scale formal validation. In patients with clinically high risk of stone recurrence (cystinuria, uncontrolled metabolic disease) or unreliable PROM assessments (neuropathic conditions) the PROM might only be used as an adjunct alongside the traditional assessments until more focussed data is available. Our study was conducted with USIQoL administered in a paper format when patients physically attended the clinic. This helped us achieve high response rates and no missing data. Sound strategies will need to be followed to achieve this when applied in other settings, such as the use of ePROM or virtual clinics. The strategies to handle missing data (imputation methods) and their full impact will need to be further explored.

There are multiple applications to our findings that would benefit different stakeholders. For the first time, the study establishes the role of PROM in urolithiasis in outpatient settings. It would improve patient engagement in their care with the implementation of patient-initiated follow-up strategies. For patients, the results would reduce frequent hospital visits and imaging. The PROM would substitute a blanket policy of traditional follow-up for all by offering alternative pathways incorporating e-PROM, nurse-led services, and possible application of artificial intelligence.

In conclusion, this study has demonstrated the usefulness of the USIQoL as an aid to outpatient management. The cut-off scores identify at risk patients with potential problems. In the lower risk patient groups, it provides a reliable tool for patient-centric evaluation and an alternative to the traditional follow-up. The prediction model is a useful triage tool that is suitable for patient initiated follow-up. The results offer framework for the large-scale validation of the model.

Declarations

References

1. Rukin, N. J., Siddiqui, Z. A., Chedgy, E. C. P., & Somani, B. K. (2017). Trends in upper tract stone disease in England: Evidence from the hospital episodes statistics database. *Urologia Internationalis, 98*(4), 391–396. https://doi.org/10.1159/000449510

2. Lang, J., Narendrula, A., El-Zawahry, A., Sindhwani, P., & Ekwenna, O. (2022). Global trends in incidence and burden of urolithiasis from 1990 to 2019: An analysis of global burden of disease study data. *European Urology Open Science, 3*(35), 37–46. https://doi.org/10.1016/j.euros.2021.10.008

3. Geraghty, R. M., Cook, P., Walker, V., & Somani, B. K. (2020). Evaluation of the economic burden of kidney stone disease in the UK: A retrospective cohort study with a mean follow-up of 19 years. *BJU International, 125*(4), 586–594. https://doi.org/10.1111/bju.14991

4. National Institute for Health and Care Excellence. Renal and ureteric stones: Assessment and management (NG 118). Retrieved March 2, 2023, from https://www.nice.org.uk/guidance/NG118

5. European Association of Urology. Urolithiasis: follow-up of urinary stones. Retrieved May 30, 2024, from https://uroweb.org/guidelines/urolithiasis/chapter/followup-of-urinary-stones

6. Lohr, K. N., & Zebrack, B. J. (2009). Using patient-reported outcomes in clinical practice: Challenges and opportunities. *Quality of Life Research, 18*, 99–107. https://doi.org/10.1007/s11136-008-9413-7

7. Kotronoulas, G., Kearney, N., Maguire, R., Harrow, A., Di Domenico, D., Croy, S., & MacGillivray, S. (2014). What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. *Journal of Clinical Oncology, 32*(14), 1480–1501. https://doi.org/10.1200/JCO.2013.53.5948

8. Kingsley, C., & Patel, S. (2017). Patient-reported outcome measures and patient-reported experience measures. *BJA Education, 17*(4), 137–144. https://doi.org/10.1093/bjaed/mkw060

9. Velikova, G., Booth, L., Smith, A. B., & BrownLynchBrownSelby, P. M. P. J. M. P. J. (2004). Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomized controlled trial. *Journal of Clinical Oncology, 22*(4), 714–724. https://doi.org/10.1200/JCO.2004.06.078

10. Assimos, D., Krambeck, A., Miller, N. L., Monga, M., Hassan-Murad, M., Nelson, C. P., Pace, K. T., Pais, V. M., Pearle, M. S., Preminger, G. M., Razvi, H., Shah, O., & Matlaga, B. R. (2016). Surgical management of stones: American Urological Association/Endourological Society Guideline PART I. *The Journal of urology, 196*(4), 1153–1160. https://doi.org/10.1016/j.juro.2016.05.090

11. Joshi, H. B., Johnson, H., Pietropaolo, A., Raja, A., Joyce, A. D., Somani, B., Philip, J., Biyani, C. S., & Pickles, T. (2022).

Urinary stones and intervention quality of life (USIQoL): Development and validation of a new core universal patient-reported outcome measure for urinary calculi. *European Urology Focus, 8*(1), 283–290. https://doi.org/10.1016/j.euf.2020.12.011

12. Crosby, R. D., Kolotkin, R. L., & Williams, G. R. (2003). Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology., 56*(5), 395–407. https://doi.org/10.1016/s0895-4356(03)00044-1

13. Ousmen, A., Touraine, C., Deliu, N., Cottone, F., Bonnetain, F., Efficace, F., Brédart, A., Mollevi, C., & Anota, A. (2018). Distribution-and anchor-based methods to determine the minimally important difference on patient-reported outcome questionnaires in oncology: A structured review. *Health and Quality of Life Outcomes, 16*(1), 228. https://doi.org/10.1186/s12955-018-1055-z

14. EuroQol Group. (1990). EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy, 16*(3), 199–208. https://doi.org/10.1016/0168-8510(90)90421-9

15. Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMC Medicine, 13*, 1. https://doi.org/10.1186/s12916-014-0241-z

16. Mcleod, L. D., Coon, C. D., Martin, S. A., Fehnel, S. E., & Hays, R. D. (2011). Interpreting patient-reported outcome results: US FDA guidance and emerging methods. *Expert Review of Pharmacoeconomics & Outcomes Research, 11*(2), 163–169. https://doi.org/10.1586/erp.11.12

17. Riley, R. D., Ensor, J., Snell, K. L. E., Harrell, F. E., Jr., Martin, G. P., Reitsma, J. B., Moons, K. G. M., Collins, G., & van Smeden, M. (2020). Calculating the sample size required for developing a clinical prediction model. *BMJ, 18*(368), m441. https://doi.org/10.1136/bmj.m441

18. Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Wiley.

19. Ruopp, M. D., Perkins, N. J., Whitcomb, B. W., & Schisterman, E. F. (2008). Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biometrical Journal, 50*(3), 419–430. https://doi.org/10.1002/bimj.200710415

20. Wright, A., Hannon, J., Hegedus, E. J., & Kavchak, A. E. (2012). Clinimetrics corner: A closer look at the minimal clinically important difference (MCID). *The Journal of Manual & Manipulative Therapy, 20*(3), 160–166. https://doi.org/10.1179/2042618612Y.0000000001

21. Bitton, A., Onega, T., Tosteson, A. N. A., & Haas, J. S. (2014). Toward a better understanding of patient-reported outcomes in clinical practice. *The American Journal of Managed Care, 20*(4), 281–283.

22. Snyder, C., & Brundage, M. (2010). Integrating patient-reported outcomes in healthcare policy, research and practice. *Expert Review of Pharmacoeconomics & Outcomes Research, 10*(4), 351–353. https://doi.org/10.1586/erp.10.21

23. Lavallee, D. C., Chenok, K. E., Love, R. M., Petersen, C., Holve, E., Segal, C. D., & Franklin, P. D. (2016). Incorporating patient-reported outcomes into health care to engage patients and enhance care. *Health Affairs (Millwood)., 35*(4), 575–582. https://doi.org/10.1377/hlthaff.2015.1362

24. Greenhalgh, J. (2009). The applications of PROs in clinical practice: What are they, do they work, and why? *Quality of Life Research, 18*(1), 115–123. https://doi.org/10.1007/s11136-008-9430-6

25. National Health Service (NHS) England. Patient Reported Outcome Measures (PROMs). Retrieved March 2, 2023, from https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/patient-reported-outcome-measures-proms

26. Lane, J. A., Donovan, J. L., Young, G. J., Davis, M., Walsh, E. I., Avery, K. N., Blazeby, J. M., Mason, M. D., Martin, R. M., Peters, T. J., & Turner, E. L. (2022). Functional and quality of life outcomes of localised prostate cancer treatments (Prostate Testing for Cancer and Treatment [ProtecT] study). *BJU International, 130*(3), 370–380. https://doi.org/10.1111/bju.15739

27. Getting It Right First Time (GIRFT). Urology outpatient transformation: A practical guide to delivery. 2022. Retrieved March 2, 2023, from https://www.gettingitrightfirsttime.co.uk/wp-content/uploads/2022/01/Urology_2022-01-12_Guidance_Outpatient-transformation.pdf

28. Cappelleri, J. C., & Bushmakin, A. G. (2014). Interpretation of patient-reported outcomes. *Statistical Methods in Medical Research, 23*(5), 460–483. https://doi.org/10.1177/0962280213476377

29. Rasmussen, A. A., Wiggers, H., Jensen, M., Berg, S. K., Rasmussen, T. B., Borregaard, B., Thrysoee, L., Thorup, C. B., Mols, R. E., Larsen, S. H., & Johnsen, S. P. (2021). Patient-reported outcomes and medication adherence in patients with heart failure. *European Heart Journal–Cardiovascular Pharmacotherapy., 7*(4), 287–295. https://doi.org/10.1093/ehjcvp/pvaa097

30. Mouelhi, Y., Jouve, E., Castelli, C., & Gentile, S. (2020). How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. *Health and Quality of Life Outcomes, 18*(1), 136. https://doi.org/10.1186/s12955-020-01344-w