Dhanya Nantha Kumar and H. B. Joshi

14 Patient-centered outcome assessments in surgical disciplines: an overview using example of the Urinary Stones and Intervention Quality of Life measure for kidney stone disease

Abstract: Surgical discipline involves rigorous assessments of outcomes, relevant to both surgeons and the recipients of care, over the short term and long term. The outcomes carry significance to other stakeholders such as the resource providers and industry partners. Patient-reported outcomes (PRO), which contribute to these assessments, are increasingly considered to be an important part of person-centered practices. In this chapter, we examine the current status of PRO assessments in the surgical field. Our focus is on applications of advanced measurement techniques and their adoption in processes relevant and useful to the various clinical stakeholders. We have divided the chapter into two parts. In the first part, we focus on the principles behind the development and validation of the Urinary Stones and Intervention Quality of Life scale, a disease- and intervention-specific PRO scale for urinary stone disease. We describe the framework, in which this new instrument has been developed, and we explore how probabilistic conjoint measurement theory has added scientific rigor to the traditional methods of PRO reliability and validity assessment. In the second part of the chapter, we provide a brief overview of the literature and examples of the probabilistic PRO measurement model's applications in the surgical branches of medicine. The current status of, and challenges surrounding, the development and application of PRO measurements of surgical outcomes are explored, with anticipations of the scope required for their wider adoption and applications, resulting in improved assessments.

Keywords: Outcome Assessments, Surgical Disciplines, USIQOL, Kidney Stones

14.1 Introduction

Patient-centered care is defined as being "respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions" (Committee on Quality of Health Care in America, 2001). It is a broad concept and forms one of the important components of quality care provision that in-

Dhanya Nantha Kumar, School of Medicine, Cardiff University, Cardiff, Wales, United Kingdom
H. B. Joshi, Department of Urology, University Hospital of Wales, Cardiff, Wales, United Kingdom

https://doi.org/10.1515/9783111036496-014

cludes person-centered approaches involving the total person and his/her own life, including caregivers. However, many factors are involved in the provision of good care, and this can be compromised by the lack of a body of well-developed theory, instruments, and evidence that substantiates the role and value of patient centered care in the broader medical context. Sometimes patient-centered care is mistakenly considered to contradict accepted standards of care. Providing and accounting for effective person-centered care must involve patients, using both qualitative and quantitative methods. These include tools such as interviews and direct observations, self-reported or performance-based measures, and more recently, measures from wearables and monitoring equipment (biomedical indicators).

Surgical disciplines manage many conditions that present in an acute and/or chronic form. The diversity of conditions demands a range of emergency and routine treatments, as well as one-off or repeated interventions. The decision to meet a patient's needs by opting for surgical interventions can pose risks. Many times, the disease can be treated in either a surgical or nonsurgical way with different risk-benefit ratios. The choice in favor of surgical management potentially involves an added degree of uncertainty due to the increased risk of morbidity and mortality during the perioperative period. Treatment options may present a clinical equipoise. Hence, patient-centered care, shared care decision making, and the understanding of patient preferences become very important.

14.1.1 Patient-reported outcome measurements (PROM): key to person-centered care

A patient-reported outcome measurement (PROM) is a report on patient's health condition that comes directly from the patient and plays an essential role in person-centered care (U.S Dept. of Health and Human Service Food and Drug administration, 2009). PROMs have been categorized as Generic or Disease-, Condition-, and Intervention-specific. Some generic measures are used as health economic tools to provide data on the quality-adjusted life years (QALYs). There can be some overlap between the aspects of health-related quality of life (HRQoL), measured by generic and disease-specific instruments. These tools have multiple applications.

In addition to the use of PROMs in randomized controlled trials to assess treatment effectiveness, there is growing interest in their use in routine HRQoL monitoring of patients and medical audits (Dept. of Health, 2010). Recent studies support the use of PROMs in clinical practice for improved shared decision-making and patient self-management (Kotronoulas et al., 2014). They have been found useful when there is a need to "identify triggers for surgery and potentially reduce the burden on services by limiting unnecessary or ineffectual procedures" (Kingsley & Patel, 2017). 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 (Velikova et al., 2004).

There is evidence for the usefulness of the PROMs in clinical practice. PROMs facilitate the detection of physical or psychological problems (Bitton et al., 2014). PROMs compare favorably with other common clinical measures in terms of reliability (Snyder & Brundage, 2010). Many national surgical bodies advocate their use to evaluate outcomes, guide routine surgical practices, and in decision making. For example, the American Urological Association (AUA) guidelines state that treatment decisions about urinary calculi should incorporate patient preferences, influenced by HRQoL impacts, rather than being limited to clinical and radiological outcomes (Penniston & Nakada, 2016).

14.1.2 General considerations behind development and application of a PROM

PROMs would contribute more consistently to improving the evidence base, supporting patient-centered care, if the measurements were more solidly grounded in science and shown to be in accordance with international standards (US FDA and Scientific Advisory Committee, 2002). The ability of PROMs to improve decision-making depends on demonstrating how they accurately capture the burden of disease or effects of treatment. PROM data should clearly indicate the meaning of small changes to the scores and when there is a need to act or decide on management plans (Bitton et al., 2014).

The methodology for the development of a PROM was established over four decades ago and has continued to evolve. It involves a multiphase approach that includes construct definition: the qualitatively informed generation of items (questions). This is followed by pilot and field testing. The final instrument is expected to satisfy demands for reliability, validity, and responsiveness. Classical Test Theory (CTT) and its focus on ordinal scores formed the main basis for demonstrating measurement quality for many years, but it is now well recognized that measurements that comply with interval scaling requirements of conjoint additivity support higher quality inferences (Terwee et al., 2018).

Rasch, in 1960, proposed a theory of measurement, producing ratio/interval scales of both stimulus and object parameters (Rasch, 1960, 1961). Andrich (1988) stated that these models, relevant to the analysis of social science data, are the same as those of the laws of physics. Their perspective was further developed by other scientists focused on paired comparisons and has been more recently been said to provide “a specifically metrological approach to human-based measurement” (Fischer & Molenaar, 1995; Andrich & Marais, 2019; Linacre, 2000; Mari et al., 2023).

In measurements modeled to be conjointly additive, the probability of a specified response (e.g., right/wrong answer, or agreeable response) being a function of the difference between each individual person’s ability or performance, and the difficulty or challenge posed by each individual item. This is an approach to mathematical model-

ing where item values are calibrated and person abilities are measured on a shared continuum quantifying the latent trait. This approach cannot guarantee but supports the development of internally valid measurements that exhibit structural invariances, independent of the sample, with findings for samples extrapolating to population characteristics and clinically meaningful differences (Pendrill, 2014; Granger, 2007). This work underpins the current application of probabilistic measurement modeling in validations of contemporary PROMS.

Criteria for judging the quality of a PROM and its validity in the clinical field have been the subject of debate. For the application of PROMs in the clinical world, COSMIN guidelines were developed to evaluate the methodological quality of studies, intended to establish the measurement properties of HRQoL scales (Hobart & Cano, 2009). When selecting a robust PROM, these guidelines advocate the use of scales developed on the basis of probabilistic measurement modeling as this increases the likelihood of covering many important steps in validity assessments. These steps include the development of data fit to a model, the demonstration of unidimensionality and obtaining satisfactory discrimination as well as evaluative properties. These steps are discussed in the next section using the example of a disease- and intervention-specific PROM for urinary calculi.

14.2 Urinary Stones and Intervention Quality of Life (USIQoL) PROM: development and validation of a disease- and intervention-specific PROM for urinary calculi

Urolithiasis is a common condition that has a global incidence of 10% (prevalence range of 2–13% across continents) amongst the general population, with 50% of patients likely to form further stones within five years (Mokkink et al., 2010). The disease caused 550,000 emergency room visits in the USA in 2009 and over 30,800 hospital admissions in England in a single recent year (Pearle et al., 2005; Hospital Episodes Statistics Data, 2014). Stone patients miss an average of 47.9 h of work per year with additional hours lost due to ambulatory care visits (Bultitude & Rees, 2012).

There are different options for managing urinary calculi with expectant, medical or interventional treatments (Saigal et al., 2005), which can be multistage and carry different risks and success rates. Urolithiasis and its treatment(s) have an adverse effect on HRQoL and can compromise all areas of patient functioning (Türk et al., 2020; Raja et al., 2016). Attempts have been made to measure HRQoL of patients with urolithiasis (Penniston & Nakada, 2016). Generic measurement scales have been used for this, but often fail to capture the clinically relevant domains (Türk et al., 2020). This has led to the introduction of the new Urinary Stones and Intervention Quality of Life (USIQoL)

questionnaire, a disease- and intervention-specific PROM that has been developed to meet the need for more relevant information.

The initial developmental work with patient interviews (62 patients and 30 family members) produced a conceptual framework and an initial long draft of the questionnaire. This generated 106 themes and 10 broad headings. These were mapped to a conceptual framework with removal of duplications to create item sets. A five-point rating scale ("not at all" to "a lot") was selected for the initial draft.

Given the five-point rating scale and the items that are reasonably on-target (such that the sample measurement mean is near the item calibration mean, and the measurement and calibration ranges of variation overlap, with no significant floor or ceiling effects), a sample size of 25 to 60 will give 99% confidence that the item estimates are within 0.5 logits of their stable value (Patel et al., 2017). A sample ranging between 200 and 400 or 500 ought then to provide four or five class intervals. The validation was performed in 2 field tests and the analysis (polytomous extended response category, partial credit model) was performed using RUMM 2030 software.

14.2.1 Field test 1

Of the total sample of 250 patients, 212 participated in this phase. The revised version of the questionnaire included 60 items. It evaluated pain using different formats for rating the frequency of mild to unbearable pain, the intensity of the worst pain, day to day as well as average pain, etc. in 10 items overall; and also addressed physical and social health (including sex life, 18 items), psychological health (6 items), work performance (8 items) and travel/holiday

Table 14.1: Example of changes to the response categories due to disordered thresholds.

Initial draft:						
Since your current stone problems began, how much have you:	Not at all	A little	Quite a bit	Very much	A lot	N/A
Had difficulty sleeping?						
Felt depressed?						
Since your current stone problems began, have your stones made you reluctant about:	Not at all	A little	Quite a bit	Very much	A lot	N/A
Making a long journey?						
Planning a holiday because you might need to use unplanned medical services?						

Table 14.1 (continued)

Final draft:					
Since your current urinary stone problems began, **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 current, urinary stone problems, **have your symptoms made you reluctant about:**	Not at all/a little		Quite a bit/a lot		N/A
Q10. Making a long journey?					

issues (3 items). Fourteen items addressed additional problems, including those arising from treatments, and others involving help from the healthcare team and family members. Finally, a single global health question was included.

The results of a traditional analysis (classical test theory) for consistency and validity showed this draft of the USIQoL to be a reliable and valid measure of impact of stones on different domains. Reliability was satisfactory, given the diagnostic purpose of the scale [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.

Further measurement scaling analyses using conjoint additivity demonstrated many limitations that were not identified by the traditional (CTT) analysis. All scales indicated good to excellent reliability, with person separation indexes (PSI) ranging between 0.62 and 0.89, given the demands for precision tolerances imposed by screening and diagnostic applications. However, almost all scales had over 60% of the items with disordered thresholds (difficulty in distinguishing between responses “quite a bit” and “very much”), necessitating change from 5 to 4 or even 2 response categories (e.g., questions evaluating ability to travel for social reasons and leisure) (Table 14.1). This controversial step of collapsing adjacent categories was taken as a preliminary and provisional effort at creating a tool, meaningful to patients (Linacre, 1994; Adams et al., 2012).

In principle, the thresholds in any scale should demonstrate response categories, representing consistently increasing levels of the construct being measured (the correct ordering of the response categories is reflected in successive thresholds). We have observed that during clinical use, having thresholds that correspond to relatable ranges in the measured construct helps in improved understanding and patient acceptability of the scale items. Item fit is evaluated using the chi-square statistics to assess that the central property of item invariance (the hierarchical ordering of the items) does not vary across the trait measured. Fit residuals demonstrate the differences between the observed and expected data for each person and item. Each scale

had items with significant fit residuals (12–60%), and residual correlations (50–90%), indicating redundancy of several items.

The removal of off-construct items, which provoked high-residual inconsistent responses, was conducted in an iterative manner, with the removal of a single item at a time, followed by reanalysis and creation of revised versions. It is important to note that this phase involves significant contributions from the clinicians and health care professionals who are experienced in the management of the target patient population. The statistical tests often result in an undecidable equipoise regarding item evaluations and so it is not always possible to select items based on analytic results alone.

The final item selection is always a multidisciplinary task. This is very important when the wider concept of validity of a PROM is to be considered. We found this to be helpful when subsequent application of the PROM in different clinical contexts was planned. The revised USIQoL included 19 questions sets divided as 5 scales of pain, social health (5 items each), physical, psychological health (4 each), and work (U.S Dept. of Health and Human Service Food and Drug administration, 2009) with 4 treatment items. This scale underwent a final validation study in a second field test.

14.2.2 Field test 2

In total, 369 of 390 patients participated in this phase. The analysis demonstrated that most of the items in the scales mapped out continua of increasing bother. 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.

We found that a revised scaling was necessary as 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). Figures 14.1–14.3 illustrate satisfactory item-threshold distribution maps of subscales. Differential item functioning (DIF) evaluates the extent to which different groups within the sample (e.g., age, anatomical site of stone [kidney or ureter], and type of intervention). This is very important clinically, especially when the target population can be very heterogeneous. The stone disease has certain clinical features (ureteric vs renal stone, with or without underlying metabolic abnormality, etc.) that can carry different QoL impact for the groups, and influence management. We evaluated all 15 questions and 3 scales against different patient subpopulations, confirming adequate performance across sample groups. This is important in the context of its wider clinical application, where valid prediction of differential behaviors across patient subpopulations is essential.

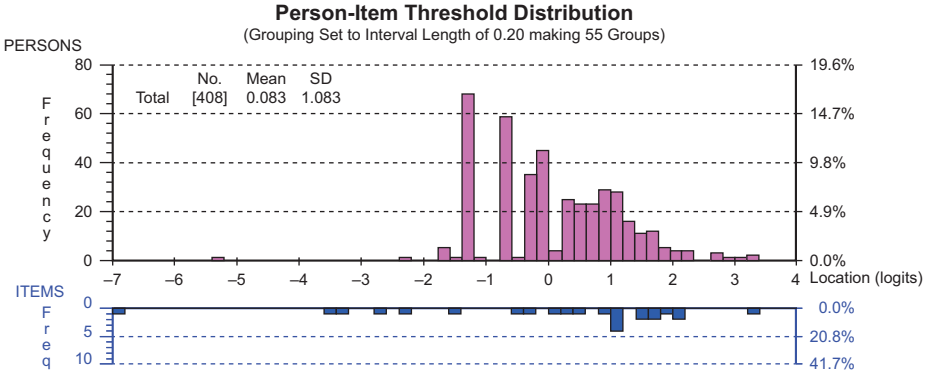

Figure 14.1: USIQoL: person–item threshold distribution – domain PPH.

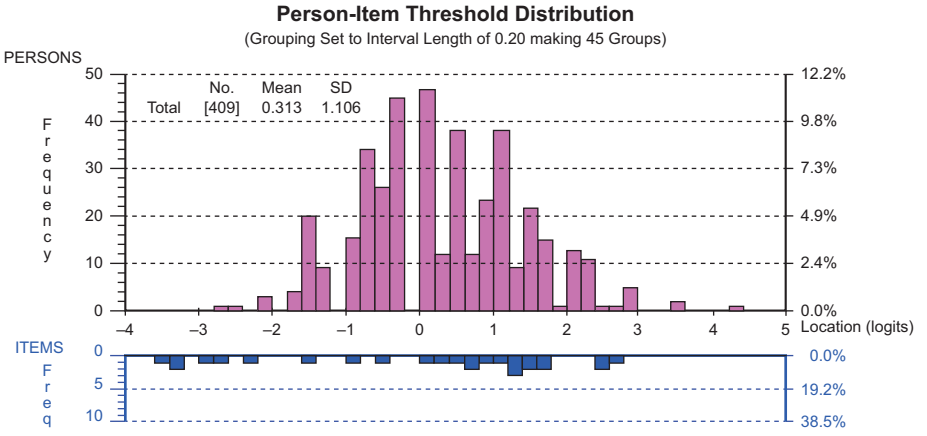

Figure 14.2: USIQoL: person–item threshold distribution – domain PSH.

Unidimensionality evaluations determine if any identifiable constructs are exhibited in the data after the main dimension has been considered. Model fit statistics indicate that all three scales of the USIQoL showed satisfactory unidimensionality. Pain, along with physical symptoms, which drives most of the clinical assessments, has clear impact. Pain, being the most complex construct to assess, was tested extensively before finalizing its format. Similarly, issues regarding work pose special data consistency problems because they are important to all stakeholders but not applicable to all patients. Conversely, the psychosocial scale is likely to be a good indicator of issues not evaluated routinely in clinical practice, and also of the longer-term impact of the condition, which could drive treatment choices. The USIQoL captures all of these dimensions well with the results quantified in a consistently interpretable, stable frame of reference.

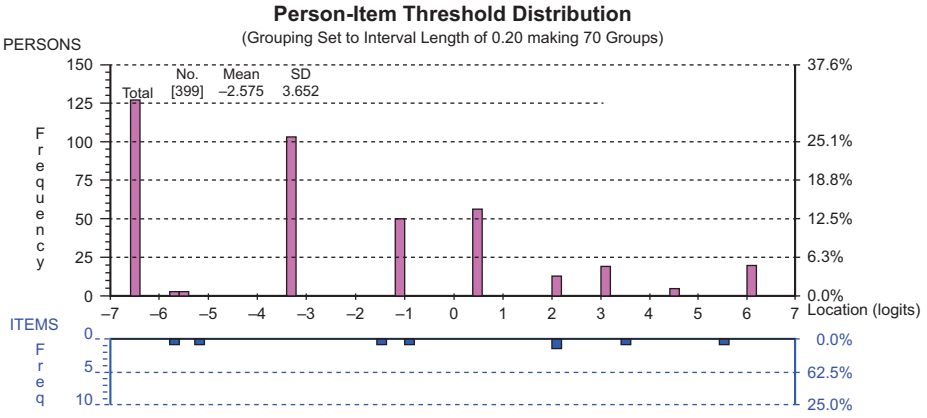

Figure 14.3: USIQoL: person–item threshold distribution – domain work.

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) (Table 14.2). Scale scores are generated by summing items and transferring to a 0–100 (logit) scale, with high scores indicating greater patient bother. It provides an internally valid measurement, demonstrably invariant, independent of the sample and with findings extrapolating to population measures of clinically meaningful differences. The final item selection in USIQoL was based on the appraisals of the analyses against clinical relevance and measurement criteria. Psychometric evaluation showed that all three scales satisfied criteria for acceptability, validity, and reliability. The logit scoring for each scale offers different scores, allowing clearer identification of the impact across different domains. The results from traditional validity assessments alone suggested that the long draft of the USIQoL satisfied most of the criteria, until probabilistic measurement demonstrated many targeting problems (e.g., disordered responses and item redundancies). This highlighted the value of conjoint probabilistic measurement to conduct item-level analyses that guide precise item selection, and rectify problems with scales.

14.2.3 Clinical application of the USIQoL: establishing validity in a wider context

It has been suggested that although robust psychometric properties of a PROM, based on consensus statements, are a precondition to use, a PROM's validity in fact lies in the sound argument that a network of empirical evidence supports the intended interpretation in a particular context (Andrich, 2013). This idea was explored by conducting a feasibility study to see if the USIQoL can be used as an aid in outpatient settings to optimize the traditional follow-up of patients with urinary calculi. Most patients

Table 14.2: USIQoL final draft of the PROM.

Urinary Stones and Intervention Quality of Life Measure

The USI-QoL – Stone Disease©

We are interested in knowing how your quality of life has been affected by your **current urinary stone problems.**

Please answer all questions on the next page, in order, by ticking the appropriate boxes.

If your feel a question is not applicable to you, please tick the 'N/A' column.

Today's date: _____

Date of birth: _____

We thank you for taking the time to complete this questionnaire

Please think about current **problems that are due to your urinary stones**

Since your current urinary stone problems, and due to urinary stone problems, **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?					
Since your current, urinary stone problems, **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 current, urinary stone problems, **with regards to the future, how much are you worried about:**					
Q9. More symptoms from your stones in the future?					
Since your current, urinary stone problems, **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 current urinary stone problems, **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?					

Table 14.2 (continued)

Urinary Stones and Intervention Quality of Life Measure **The USI-QoL – Stone Disease©**					
Since your current urinary stone problems, **how much have you found yourself having problems with:**					
Q12. Having to take medication (painkillers, preventative treatment etc.)?					
Q13. Increasing your water intake?					
Work					
Please mark 'Not applicable' (N/A) if currently not working (paid employment).					
Since your current urinary stone problems **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?					

with urolithiasis undergo long-term follow-up involving regular clinic review and imaging to prevent or identify possible complications early. This is resource-intensive, involves exposure to ionizing radiation, and is not without diagnostic limitations. Furthermore, there are wide variations in practices. Deciding the optimal frequency and duration of follow-up for stones is a longstanding problem with little evidence base and alternatives. The National Institute of Clinical Excellence (NICE) in the UK indicated that currently no recommendations can be made regarding follow-up and that more research is needed (Hawkins et al., 2018).

The important question in need of answering when a patient with urolithiasis attends a clinic is whether the stone(s) need an intervention to treat or can be monitored. To this end, it would be important to know if the adoption of the USIQoL as a monitoring tool can assist clinical -making. This would be suitable if the results correlate well with those of traditional follow-up methods, or with outpatient review involving consultation and imaging. In the latest Urology Outpatient Transformation guide in the UK, "personalized follow up – patient-initiated follow-up" and "using remote monitoring" were highlighted as two key components within the scope of improved PROM-based follow-up (National Institute for Health and Care Excellence, 2023). Following the COVID pandemic, there are pressures for changes to outpatient practices and increased acceptance of alternative methods of follow-up (National Institute for Health and Care Excellence, 2023).

Hence a feasibility study was conducted to establish the validity of the USIQoL in a wider clinical context (Getting It Right First Time (GIRFT), 2023), with three objectives:

1) To assess the validity of the USIQoL as an outcome measure in the outpatient setting by establishing its correlation with the traditional follow-up (current standard of care).
2) To develop valid USIQoL cutoff scores that can reliably differentiate between patients who need active treatment against those who do not and facilitate a follow-up strategy, including remote methods.
3) To define 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.

Initially, the USIQoL-based decision model was developed using existing data. Subsequently, a prospective, single-blind validation of the model for outpatients was conducted. For subjective measures, in general, including the application of the PROMs, the FDA recommends different types of anchors as external criteria, approximating truth, to generate relevant thresholds for meaningful within-patient change. These recommended anchors are

1) well-established clinical outcomes (intervention or not in our case);
2) global impressions of change in stone-related symptoms; and
3) static – current-state global impression of severity (EQ-5D PROM, in this case).

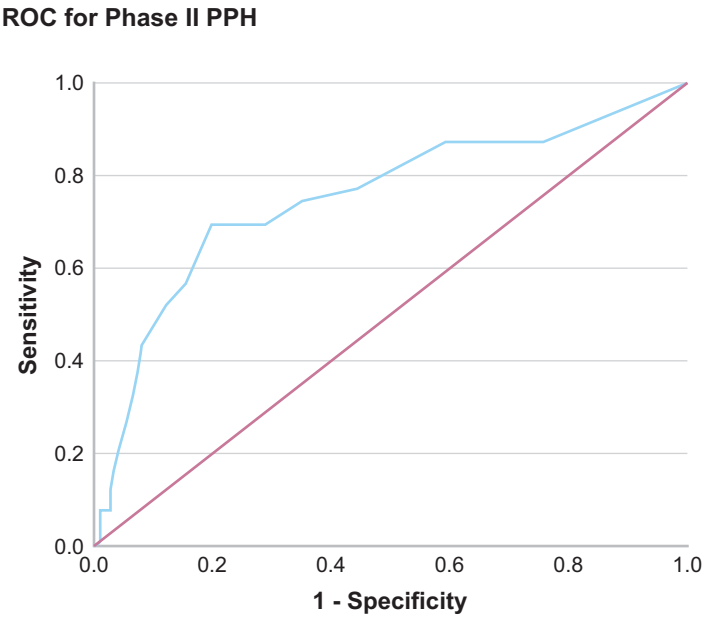

Figure 14.4: ROC curves for PPH (Pain and Physical health) domain – Phase II.

For the purposes of this study, USIQoL measurements from the two major domain scales (PPH and PSH) were considered. The study assessed correlation between the USIQoL measurements and the outcomes listed above. The study helped to validate USIQoL cutoff standards to discriminate between patients' needs to intervene or not. Analysis involved binomial logistic regression (BLR) and receiver operating characteristic (ROC) curves.

Data from 455 patients showed that the relationship between USIQoL scores (Pain and Physical Health, PPH and Psycho-Social Health, PSH domains) and clinical outcomes were statistically significant [estimated odds: PPH 1.24, $p < 0.001$, 95% CI 1.13–1.36; PSH 1.179, $p < 0.001$, 95% CI 1.12–1.33]. The ROC values were >0.75 when an Area under curve (AUC) of 0.7 to 0.8 is considered acceptable (Jarvis et al., 2023) (Figures 14.4 and 14.5). This demonstrated satisfactory ability of the model to differentiate between the two clinical outcomes. The optimum cutoff measurements were found to be 9 (PPH) and 11 (PSH), based on the Youden index.

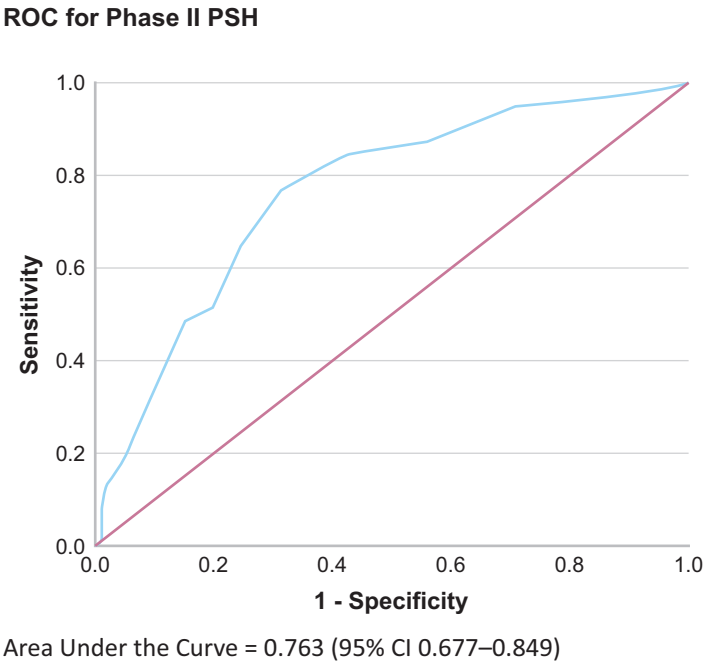

Figure 14.5: ROC curve for PSH (psycho-social health) domain – Phase II.

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 standardized. It is well known that MCID is a complex concept with multiple facets and variable results, based on the methods used. Combinations of anchor- and distribution-based methods were used to give the best estimates. The Minimally Clinically Important Dif-

ference (MCID) for the domains scores was 3–4 points. The model demonstrated satisfactory sensitivity (0.90) and specificity (0.46). Using this, it was clear that the odds of patients expressing symptoms and then needing full clinical evaluation with imaging and active intervention, increased with the increasing USIQoL scores. The results confirmed good correlation and one-dimensionality between the PPH and PSH domains.

Thus, the feasibility study demonstrated good correlation between the PROM and the clinical outcomes, making it a valid aid for outpatients. The cutoff scores identify patients at risk. It provides a reliable tool for patient-centric evaluation and an alternative to the long-term traditional follow-up policy, and established validity of USIQoL in a wider context.

14.3 PROMs in surgical disciplines: overview of the literature

We explored the current status of the PROMs in surgical disciplines, with a focus on application of the metrologically oriented measurement theory. Although the formal systematic review is out of the scope of this chapter, we have worked out the broad trends and key messages using examples from the literature. The implications are discussed in more detail in the subsequent section.

For the search, "patient reported outcomes," "surgery," "applications," "outcome metrology," "Rasch analysis," "conjoint measurement," and "decision making" were the key words used. The search, covering over 3 decades, resulted in over 18,000 articles with PROM and over 8,000 articles with Rasch key words. The results covered studies with significant heterogeneity. These largely reported on the developmental work on PROMs, or comparative studies, when applied to a cohort of patients in a single or multicenter study.

At the micro-level, PROMs facilitate the detection of clinical problems and adherence to treatments (Bitton et al., 2014). Real-time access to the PROM data helps clinicians prioritize topics for discussion at review and improves patient–clinician communication (Rasmussen et al., 2021). At the meso-level, PROM data can help in comparative effectiveness research and evaluation of the impact of interventions (Lavallee et al., 2016). There are four main mechanisms used internationally for the routine collection and aggregation of PRO information (Greenhalgh, 2009). Some of these have been exclusively used in the area of surgical practices:

A) Pre- and post-procedure data collection from patients undergoing selected elective surgeries to assess hospital performance (e.g., the National NHS PROMs program): Four surgical procedures were initially chosen to be included in the national PROMs program (2009 on), mandated in the NHS Outcomes, and included total hip replacement, total knee replacement, varicose veins (until 2017),

and groin hernia surgery (until 2017). The main aim was to benchmark procedural outcomes across different trusts (Williams et al., 2016).

B) Computer-assisted testing using banks of questions that capture generic patient-reported outcomes, common across a number of chronic conditions (e.g., the US-based Patient Reported Outcomes Measurement Information System (PROMIS) initiative): This is aimed at providing patient-level data using pre-prepared question banks covering different domains (Coles, 2010).
C) Inclusion of PROMs within disease-specific clinical registries (e.g., the Swedish Healthcare Quality Registries).
D) International initiatives to develop standard outcome measurement sets, including PROMs, to foster international benchmarking (e.g., International Consortium for Health Outcomes Measurement).

This literature thus demonstrates the current wide-ranging applications of PROMs.

14.3.1 PROMS and evaluation using additive conjoint measurement modeling techniques

14.3.1.1 Development of new PROMS

Many new PROMs covering different surgical disciplines have been developed using conjoint measurement theory and modeling over the last 15 years, with many employed in evaluating clinical trial outcomes (PROMIS®; Joshi et al., 2022; Pesudovs et al., 2004). A new 20-item Quality of Life Impact of Refractive Correction (QIRC) questionnaire, which quantifies the QOL of people with refractive correction by spectacles, contact lenses, and refractive surgery in the prepresbyopic age group, was developed by Pesudovs et al. in 2004 and has been shown to have broad applicability for cross-sectional and outcomes research (Joshi et al., 2022). Similarly, the BREAST-Q is a PROM used to assess the unique outcomes of breast surgery patients that was developed in 2009 using conjoint measurement modeling; it is composed of three procedure-specific modules: augmentation, reduction, and reconstruction and has been used in multiple studies along with linguistic validations (Pesudovs et al., 2004).

14.3.1.2 Reevaluation of existing PROMs

Over the last two decades, the properties of existing PROMs have been developer-evaluated. Surgical disciplines such as orthopedics and ophthalmology have been at the forefront in these efforts. The results are mixed and have repeatedly demonstrated and substantiated the importance of adopting rigorous measurement modeling theory and

practice. Many existing PROMS have been found to have problems with suboptimal targeting, item fit, disordered thresholds, and a lack of meaningful and interpretable unidimensionality. This has raised questions about the validity of the measurements and the results generated using these PROMs.

The National Eye Institute Refractive Error Quality of Life instrument (NEI-RQL-42) is a commonly used questionnaire that seeks to measure refractive error-related quality of life (QoL). In light of the results produced by conjoint measurement modeling, the authors stated that NEI-RQL-42 questionnaire is deficient for all psychometric properties tested and advise clinicians or researchers to consider other questionnaires that have been more rigorously developed to meet standard psychometric properties (Pusic et al., 2009). Another study was conducted in patients with prostate cancer, undergoing radical surgical treatment. The outcomes from the surgery were monitored using the patient-reported outcome measure: Symptom Tracking and Reporting tool (STAR) (Alinden et al., 2011). This tool has four domains, which investigates sexual function, urinary function, bowel function, and overall quality of life. The study showed that urinary and sexual function scales produced inconsistent observations, insufficient to the task of measurement. The study concluded that further evaluation needs to be carried out to determine the suitability of this PROM.

A study of Patient- and Parent-Reported Outcome Measures in the International Consortium for Health Outcomes Measurement Standard Set for Cleft Lip and Palate came to similar conclusions (Protopapa et al., 2020). The study concluded that the NOSE and COHIP-OSS questionnaires were inaccurate, and that the CLEFT-Q questionnaire did not cover facial function and speech domains sufficiently. The study concluded that the PROMs used for cleft care do not satisfy the need for quantitative measurements of the outcomes produced.

Re-evaluations have also been conducted for many short-form versions of existing questionnaires (Apon et al., 2021; Multanen et al., 2020). The reviews show that, in spite of many advances over four decades, it is still challenging to select reliable tools (Lundström & Pesudovs, 2009). Of the 315 generic and condition-specific PROMs published between the 1980s and 2019, the vast majority were related to musculoskeletal conditions, with other patient-related outcomes related to cancer, gastrointestinal, mental health, and many other conditions. Of the 315 studies identified, 270 (85.7%) had been used in subsequent studies, and 45 did not have any online evidence of applications, following validation.

14.3.2 Challenges in using PROMs in clinical practice

There are multiple challenges in the implementation of PROMs and these encompass different aspects of PROM usage. The challenges can be identified at different stages:

A. Development
   1) Scientifically rigorous modeling is essential when developing measuring tools that are valid and reliable. In this regard, there has been considerable confusion within the scientific and the clinical communities regarding the viability of meaningful quantification, and the associated terminologies used in the field of PROM development and validation. This has had deleterious impacts on the interpretation and adoption of PROMs by clinical teams (Churruca et al., 2021; Derriennic et al., 2019; Hobart et al., 2007). Efforts undertaken by different agencies, such as COSMIN, are intended to standardize the nomenclature (Hobart et al., 2010).
   2) Establishing metrological standards is essential for maximizing the value of the widespread use of a PROM. COSMIN standards for the validity of measurements include criteria that can be met only if the PROM has been developed using additive conjoint measurement modeling and so demonstrates validity in a wider context than that available using ordinal measurement methods (Churruca et al., 2021; Derriennic et al., 2019; Hobart et al., 2007, 2010; Prinsen et al., 2018; Hawkins et al., 2018; Snyder & Brundage, 2010; Fisher, 2023; Allen & Pak, 2023; Massof & Bradley, 2023). This is a desirable long-term strategy that needs to be endorsed by all stakeholders. The data from such work would establish a robust evidence base for patient-centered practices; with ongoing application of the insights of the new institutional economics, such standards may one day be legally enforced, with significant implications for health care markets (Snyder & Brundage, 2010).

B. Clinical applications
   1) The selection of instruments appropriate for a given range of conditions and interventions can be challenging. There is a need for standardized assessments of the psychometric properties and validities of PROMs so that information provided is sensitive, relevant, and specific to various contexts. Provisionally resolving the tensions between standardization and personalization (Fisher, 2023; Allen & Pak, 2023; Massof & Bradley, 2023; Lipscomb et al., 2007; Mallinson, 2024) via meaningful scaling and individualized reporting is essential for generalized improvements in deciding the superiority or inferiority of surgical approaches or policies (Massof & Bradley, 2023).
   2) There is a need to improve the comparability of PRO measurements and data across different healthcare settings, countries, and cultures, which poses challenging but not intractable problems (Lipscomb et al., 2007; Mallinson, 2024).
   3) Patient and stakeholder engagement with diagnostic, treatment, and follow-up processes can be facilitated by improved measurement, as high-quality, actionable information provided confidentially via easy-to-use electronic interfaces may work to increase response rates in contexts involving the need to complete the PROMS on a repeated basis (for pre and post intervention assessments) (Massof & Bradley, 2023). Concerns expressed by clinicians have

included the time and effort involved in data collection and analysis, and the provision of adequate resources to collect the data and its analysis. Investment is required when establishing platforms for data collection and optimizing the flow and analysis of the data but may pay remarkable returns when systems are well-designed (Snyder & Brundage, 2010).

4) Data needs to be presented in forms usable to all the stakeholders at all levels (Snyder & Brundage, 2010) with clear information on what, if anything, small changes to the scores actually mean clinically (Fisher, 2023; Allen & Pak, 2023); anything less can risk leading to clinician disengagement (Bitton et al., 2014). Measurements should provide information on quality indicators; PROM data has been used to this effect in the UK in national audits covering the index orthopedic procedures (Williams et al., 2016).
5) Studies have documented limitations of existing PROMs without the application of probabilistic conjoint modeling (Pusic et al., 2009; Alinden et al., 2011; Protopapa et al., 2020). Fresh perspectives on standard setting are needed to revise the old measurement scales or develop new ones.

14.3.3 Opportunities for PROM applications

Surgical disciplines continue to evolve as minimally invasive and robotic techniques increasingly complement patient-centered practices. This offers opportunities for incorporating PROMs at every level of practice.

Micro: These are at the clinician–patient interactions level, where measurement reports are individualized to specific patients in the course of care, and to specific clinicians in the course of clinical management (Sul, 2024; Chien et al., 2009). There is evidence of benefits from these processes (Wright et al., 1980) as they contribute to improved patient counselling, ahead of interventions and the development of appropriate patient information leaflets. However, it is yet not established if the individual health status outcomes are consistently improved or not.

Meso: At the meso-level, PROMs are widely shown to be effective (Derriennic et al., 2019; Hobart et al., 2007; Fisher, 2023; Allen & Pak, 2023). This applies to the comparative effectiveness research used to investigate benefits of different treatment and surgical interventions. PRO data used in the registries helped quality improvement programs (NHS UK PROMS programs) and has improved understanding of the variations in care, costs, and outcomes. One of the major applications of PROMs is in the adoption of Value-Based Health Care (VBHC) (Chien et al., 2018). Although healthcare funding varies between different settings worldwide, there is a gradual shift from fee-for-service to the more VBHC. It aims to reduce unnecessary variations and costs in the practices. Person-centered data would provide valuable support to such programs.

Macro: There is growing interest in the development of predictive theories and explanatory models capable of independently validating the construct measured (Squitieri et al., 2017; Melin et al., 2021, 2023). Work in this area and in the programs advanced by groups such as the International Consortium for Health Outcomes (ICHOM) will help to foster international benchmarking.

More work needs to be done to advance the adoption of mass-customizable PROMs in a uniform and structured fashion. Use of measurements based on probabilistic, additive conjoint modeling analysis, with well-defined MCIDs and validity in the wider contexts, can plausibly be expected to result in new levels of utility, effectiveness, and efficiency. Standards will need to be established for unit definitions, laboratory accreditation, conformity assessment, and quality-assured traceability (Chan et al., 2015). National and international standards bodies and specialty organizations will need to focus complex cross-disciplinary initiatives on the demands of practice to devise and set the necessary guidelines. Health care insurers, funders, providers, regulators, and advocacy groups will need to collaborate in new ways to provide the necessary support and infrastructure.

Clear and interpretable standards of these kinds will support the creation of an entirely new class of quality improvement programs. It will offer opportunities for the development of systems capable of guiding systematic responses to PROMS feedback. Improvements to health information systems and technology (Jeckelmann et al., 2023) will address barriers to data collection and workload management by implementing computer-adaptive and AI measurement strategies and integrating PROM data in health records (National Institute for Health and Care Excellence, 2023). Close attention to envisioning, planning, and resourcing the needed broad scope for training and professionally developing clinicians and associated staff will pay significant substantive and financial returns as we achieve the timely dissemination of more relevant and meaningful information.

References

Adams, R. J., Wu, M., & Wilson, M. (2012). The Rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement*, *72*(4), 547–573.

Alinden, C. M., Skiadaresi, E., Moore, J., & Pesudovs, K. (2011). Subscale Assessment of the NEI-RQL-42 Questionnaire with Rasch Analysis. *Clinical and Epidemiologic Research*, *52*, 8.

Allen, D. D., & Pak, S. (2023). Improving clinical practice with person-centered outcome measurement. In W. P. Fisher Jr. & S. J. Cano (Eds.). *Person centered outcome metrology* (pp. 53–105). Springer.

Andrich, D. (1988). *Sage university paper series on quantitative applications in the social sciences*. Vol. series no. 07–068, Rasch models for measurement. Sage Publications.

Andrich, D. (2013). An expanded derivation of the threshold structure of the polytomous Rasch model that dispels any 'threshold disorder controversy. *Educational and Psychological Measurement*, *73*(1), 78–124.

Andrich, D., & Marais, I. (2019). *A course in Rasch measurement theory: Measuring in the educational, social, and health sciences*. Springer.

Apon, I., van Leeuwen, N., Allori, A. C., Rogers-Vizena, C. R., Koudstaal, M. J., Wolvius, E. B., Cano, S. J., Klassen, A. F., & Versnel, S. L. (2021). Rasch analysis of patient- and parent-reported outcome measures in the international consortium for health outcomes measurement standard set for cleft lip and palate. *Value in Health*, *24*(3), 404–412. doi:10.1016/j.jval.2020.10.019

Bitton, A., Onega, T., Tosteson, A. N. A., & Haas, J. S. (2014). Toward a better understanding of patient-reported outcomes in clinical practice. *American Journal of Managed Care*, *20*(4), 281–283. Accessed March 2, 2023, https://pubmed.ncbi.nlm.nih.gov/24884859/

Bultitude, M., & Rees, J. (2012). Management of renal colic. *BMJ*, *345*, e5499.

Chan, T. L., Perlmutter, M. S., Andrews, M., Sunness, J. S., Goldstein, J. E., & Massof, R. W., Low Vision Research Network (LOVRNET) Study Group. (2015). Equating visual function scales to facilitate reporting of Medicare functional g-code severity/complexity modifiers for low-vision patients. *Archives of Physical Medicine and Rehabilitation*, *96*(10), 1859–1865.

Chien, T. W., Chang, Y., Wen, K. S., & Uen, Y. H. (2018). Using graphical representations to enhance the quality-of-care for colorectal cancer patients. *European Journal of Cancer Care*, *27*(1), e12591.

Chien, T.-W., Wang, W.-C., Wang, H.-Y., & Lin, H.-J. (2009). Online assessment of patients' views on hospital performances using Rasch model's KIDMAP diagram. *BMC Health Services Research*, *9*, 135.

Churruca, K., et al. (2021). Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues. *Health Expect*, *24*(4), 1015–1024. doi:10.1111/hex.13254

Coles, J. (2010). PROMs risk adjustment methodology guide for general surgery and orthopaedic procedures. Hertfordshire, UK: Northgate Information Solutions (UK) Limited, 54 pp.

Committee on Quality of Health Care in America, (2001). Improving the 21st-century health care system. In Crossing the quality chasm: A new health system for the 21st century (ed., pp. 39–60). Washington, DC: National Academy Press.

Dept of Health. (2010). *Equity and Excellence: Liberating the NHS*. London Dept of health.

Derriennic, J., Nabbe, P., Barais, M., Lalande, S., Le Goff, D., Pourtau, T., Penpennic, B., & Le Reste, J. Y. Quality of primary care from the patient's point of view. A systematic review. Preprint, Universite de Bretagne Occidentale Faculte de Medecine et Des Sciences de la Sante de Brest, 2019, September 9. https://doi.org/10.21203/rs.2.14226/v1

Fischer, G. H., & Molenaar, I. (1995). *Rasch models: Foundations, recent developments, and applications*. Springer-Verlag.

Fisher, W. P., Jr. (2023). Measurement systems, brilliant results, and brilliant processes in healthcare. In W. P. Fisher Jr. & S. Cano (Eds.). *Person-centered outcome metrology* (pp. 357–396). Springer.

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

Granger, C. (2007). Rasch Analysis is important to understand and use for measurement. *Rasch Measurement Transactions*, *21*(3), 1122–1123.

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. doi:10.1007/s11136-008-9430-611

Hawkins, M., Elsworth, G., & Osborne, R. (2018). Application of validity theory and methodology to patient-reported outcome measures (PROMs): Building an argument for validity. *Quality of Life Research*, *27*, 1695–1710.

Hawkins, M., Elsworth, G., & Osborne, R. (2018). Application of validity theory and methodology to patient-reported outcome measures (PROMs): Building an argument for validity. *Quality of Life Research*, *27*, 1695–1710.

Hobart, J., & Cano, S. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: The role of new psychometric methods. *Health Technol Assess*, *13*(12), 1–200.

Hobart, J. C., Cano, S. J., & Thompson, A. J. (2010). Effect sizes can be misleading: Is it time to change the way we measure change? *Journal of Neurology, Neurosurgery, & Psychiatry*, *81*, 1044–1048.

Hobart, J. C., Cano, S. J., Zajicek, J. P., & Thompson, A. J. (2007). Rating scales as outcome measures for clinical trials in neurology: Problems, solutions, and recommendations. *The Lancet Neurology*, *6*, 1094–1105.

Hospital Episodes Statistics. (2014). National Health Service, http://www.hscic.gov.uk/hes.
Jarvis, R., Pallman, P., & Joshi, H. (2023). Is remote follow-up using patient reported outcome measure (PROM) feasible in patients with urolithiasis? The results of the first prospective feasibility study using urinary stones and intervention quality of life (USIQoL) core measure. *European Urology*.
Jeckelmann, B., & Edelmaier, R. (2023). Metrological infrastructure. *De Gruyter Series in Measurement Sciences*. De Gruyter Oldenbourg.
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. doi:10.1016/j.euf.2020.12.011, Epub 2021 Jan 8. PMID: 33423970
Kingsley, C., & Patel, S. (2017). Patient-reported outcome measures and patient-reported experience measures. *BJA Education*, *17*(4), 137–144. doi:10.1093/bjaed/mkw060
Kotronoulas, G., Kearney, N., Maguire, R., et al. (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. doi:10.1200/JCO.2013.53.5948
Lavallee, D. C., Chenok, K. E., Love, R. M., et al. (2016). Incorporating patient-reported outcomes into health care to engage patients and enhance care. *Health Aff (Millwood)*, *35*(4), 575–582. doi:10.1377/hlthaff.2015.1362
Linacre, J. M. (1994). Sample size and item calibration [or person measure] stability. *Rasch Measurement Transactions*, *7*(4), 328.
Linacre, J. M. (2000). Almost the Zermelo model? *Rasch Measurement Transactions*, *14*(2), 754. http://www.rasch.org/rmt/rmt142k.htm
Lipscomb, J., Gotay, C. C., & Snyder, C. F. (2007). Patient-reported outcomes in cancer: A review of recent research and policy initiatives. *CA: A Cancer Journal for Clinicians*, *57*, 278–300.
Lundström, M., & Pesudovs, K. (2009). Catquest-9SF patient outcomes questionnaire: Nine-item short-form Rasch-scaled revision of the Catquest questionnaire. *Journal of Cataract & Refractive Surgery* , *35*(3), 504–513. doi:10.1016/j.jcrs.2008.11.038, PMID: 19251145.
Mallinson, T. (2024). Extending the justice-oriented, anti-racist framework for validity testing to the application of measurement theory in re(developing) rehabilitation assessments. In W. P. Fisher Jr. & L. Pendrill (Eds.). *Models, measurement, and metrology extending the SI*. De Gruyter.
Mari, L., Wilson, M., & Maul, A. (2023). *Measurement across the sciences: Developing a shared concept system for measurement*. 2nd ed, Springer Series in Measurement Science and Technology, Springer.
Massof, R. W., & Bradley, C. (2023). An adaptive strategy for measuring patient-reported outcomes. In W. P. Fisher Jr. & S. J. Cano (Eds.). *Person-centered outcome metrology* (pp. 107–150). Springer.
Melin, J., Cano, S., & Pendrill, L. (2021). The role of entropy in construct specification equations (CSE) to improve the validity of memory tests. *Entropy*, *23*(2), 212.
Melin, J., Cano, S. J., Gillman, A., Marquis, S., Flöel, A., Göschel, L., & Pendrill, L. R. (2023). Traceability and comparability through crosswalks with the NeuroMET memory metric. *Scientific Reports*, *13*(1), 5179.
Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & De Vet, H. C. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, *19*(4), 539–549. doi:10.1007/s11136-010-9606-8, Epub 2010 Feb 19. PMID: 20169472; PMCID: PMC2852520.
Multanen, J., Ylinen, J., Karjalainen, T., et al. (2020). Structural validity of the Boston Carpal Tunnel Questionnaire and its short version, the 6-Item CTS symptoms scale: A Rasch analysis one year after surgery. *BMC Musculoskeletal Disorders*, *21*, 609. https://doi.org/10.1186/s12891-020-03626-2
National Institute for Health and Care Excellence. (2023). Renal and ureteric stones: assessment and management (NG 118). Accessed March 2, https://www.nice.org.uk/guidance/NG118.

Patel, N., Brown, R. D., Sarkissian, C., De, S., & Monga, M. (2017). Quality of life and urolithiasis: The patient – Reported outcomes measurement information system (PROMIS). *International Brazilian Journal of Urology*, *43*, 880–886.

Pearle, M. S., Calhoun, E. A., & Curhan, G. (2005). Urologic diseases in America project: Urolithiasis. *The Journal of Urology*, *173*, 848.

Pendrill, L. R. (2014). Man as a measurement instrument [Special Feature]. *NCSLi Measure: The Journal of Measurement Science*, *9*(4), 22–33.

Penniston, K. L., & Nakada, S. Y. (2016). Treatment expectations and health-related quality of life in stone formers. *Current Opinion in Urology*, *26*, 50–55.

Penniston, K. L., & Nakada, S. Y. (2016). Treatment expectations and health-related quality of life in stone formers. *Current Opinion in Urology*, *26*, 50–55.

Pesudovs, K., Garamendi, E., & Elliott, D. B. (2004 Oct). The Quality of life impact of refractive correction (QIRC) questionnaire: Development and validation. *Optometry and Vision Science*, *81*(10), 769–777. doi:10.1097/00006324-200410000-00009, PMID: 15557851.

Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*, 1147–1157.

PROMIS® (Patient-Reported Outcomes Measurement Information System®): US Department of health and human services: https://www.healthmeasures.net/explore-measurement-systems/promis

Protopapa, E., van der Meulen, J., Moore, C. M., & Smith, S. C. (2020). Assessment of a patient-reported outcome measure in men with prostate cancer who had radical surgery: A Rasch analysis. BMJ Open, 10(11), e035436. doi:10.1136/bmjopen-2019-035436

Pusic, A. L., Klassen, A. F., Scott, A. M., Klok, J. A., Cordeiro, P. G., & Cano, S. J. (2009). Development of a new patient-reported outcome measure for breast surgery: The BREAST-Q. *Plastic and Reconstructive Surgery*, *124*(2), 345–353. 19644246.

Raja, A., Hekmati, Z., & Joshi, H. B. (2016). How do urinary calculi influence health-related quality of life and patient treatment preference: A systematic review. *Journal of Endourology*, *30*, 727–743.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. (Reprint, with Foreword and Afterword by B. D. Wright, Chicago: University of Chicago Press, 1980). Danmarks Paedogogiske Institut. Danmarks Paedogogiske Institut.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Ed.), *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability: Volume IV: Contributions to biology and problems of medicine* (pp. 321–333). http://www.rasch.org/memo1960.pdf. University of California Press

Rasmussen, A. A., Wiggers, H., Jensen, M., et al. (2021). Patient-reported outcomes and medication adherence in patients with heart failure. *European Heart Journal –Cardiovascular Pharmacotherapy*, *7*(4), 287–295. doi:10.1093/ehjcvp/pvaa097

Saigal, C. S., Joyce, G., Timilsina, A. R., et al (2005). Direct and indirect costs of nephrolithiasis in an employed population: Opportunity for disease management? *Kidney International*, *68*, 1808–14.

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

Squitieri, L., Bozic, K. J., & Pusic, A. L. (2017). The role of patient-reported outcome measures in value-based payment reform. *Value Health* Jun, *20*(6), 834–836. doi:10.1016/j.jval.2017.02.003, Epub 2017 Mar 22. PMID: 28577702; PMCID: PMC5735998.

Sul, D. (2024). Situating culturally specific assessment development within the disjuncture-response dialectic. In W. P. Fisher Jr. & L. Pendrill (Eds.). *Models, measurement, and metrology extending the SI*. De Gruyter.

Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., et al. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research*, *27*(5), 1159–1170.

Türk, C., Neisius, A., Petrik, A., Seitz, C., Skolarikos, A., & Thomas, K. (2020). European Urological Association urolithiasis guidelines http://uroweb.org/guideline/urolithiasis/2020

U.S Dept of Health and Human Service Food and Drug administration.(2009). *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.

US FDA and Scientific Advisory Committee of the Medical Outcomes Trust: Assessing health status and quality of life instruments: attributes and review criteria. Quality Research. (2002). The ability of the PROMs to improve decision-making relies on them accurately capturing the burden of disease or treatment.

Velikova, G., Booth, L., Smith, A. B., et al. (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. doi:10.1200/JCO.2004.06.078

Williams, K., Sansoni, J., Morris, D., Grootemaat, P., & Thompson, C. 2016. Patient-reported outcome measures: Literature review Australian commission on safety and quality in healthcare.

Wright, B. D., Mead, R. J., & Ludlow, L. H. (1980). KIDMAP: person-by-item interaction mapping (MESA Memorandum #29). Chicago:. MESA Press [http://www.rasch.org/memo29.pdf].