

Outcome Instruments for Patellofemoral Arthroplasty

Elizabeth W. Paxton, MA; and Donald C. Fithian, MD

Although there are numerous patient outcome instruments available, the most reliable and valid instruments for evaluating patient outcomes after patellofemoral arthroplasty have not been identified. In this article, we review and evaluate the psychometric properties and practical considerations of administering general health instruments (Medical Outcomes Study (MOS) Short Form-36 and Short Form-12), knee scales (Knee Society Clinical Rating System, Knee Outcome Survey, International Knee Documentation Committee form, Knee Injury and Osteoarthritis Outcome Score) and a disease specific scale (Western Ontario and McMaster Universities Osteoarthritis Index) for patellofemoral arthroplasty outcome assessment. Based on our review of the literature, we recommend the Short Form-36 and Knee Injury and Osteoarthritis Outcome Score for evaluation of patellofemoral arthroplasty outcomes and provide recommendations for implementation of these instruments in a clinical setting.

Various instruments have been used to describe outcomes of patellofemoral arthroplasty (PFA), including the Knee Society Score,¹⁴ the Hungerford and Kenna knee rating scale and modifications of this scale,^{1,15} the Lysholm scale,² the Mansat scoring system,⁷ and subjective patient-reported improvement and satisfaction ratings.^{2,5} To date, the most reliable and valid instruments for evaluating patient outcomes following PFA have not been identified. Identification of practical instruments with strong psychometric properties would improve interpretation and comparison of results across studies and allow both clinicians and researchers to more accurately evaluate patient outcomes following PFA.

In selecting the most ideal instrument for the study of outcomes in PFA, we must keep in mind that these patients and the outcomes of interest to them are similar to, but not

necessarily the same as, those in other patellofemoral disorders or in total knee arthroplasty (TKA). Isolated patellofemoral arthritis may occur in patients with a history of patellofemoral instability or pain, or as a sequela of surgery of the patellofemoral joint. In either case, a patient may be considered a candidate for PFA. Cohorts of patients having PFA tend to be younger than the population of patients having TKA. It therefore makes sense to look at outcome tools developed for these two patient populations in seeking a tool for use in studying outcomes of PFA.

The instruments that have been used to measure outcomes of patellofemoral disorders, knee arthritis and TKA vary in reliability, validity, length of time to complete, complexity of scoring mechanisms, and method of administration. It clearly would be useful to standardize reporting of patient outcomes in order to improve the base of evidence on which we make treatment decisions. To help clinicians and researchers who might wish to consider using one or more of the tools for assessing patient outcomes, it would be useful to compare the psychometric and practical considerations of administering general health instruments [Medical Outcomes Study (MOS) Short Form-36 (OSSF-36) and SF-12], knee scales (Knee Society Clinical Rating System, Knee Outcome Survey, International Knee Documentation Committee form, Knee Injury and Osteoarthritis Outcome Score), and a disease specific scale (Western Ontario and McMaster Universities Osteoarthritis Index) for PFA outcome assessment. In this article, we evaluate the psychiatric properties and practical considerations of administering these instruments. Based on our review of the available measures, we identify the optimal instruments for evaluating PFA outcomes and provide recommendations for implementation of these tools in a clinical setting.

Outcome Instrument Review Criteria

Three psychometric properties are critical for comparing potential outcome tools: instrument validity, reliability, and responsiveness. Validity refers to whether a tool actually evaluates what it intends to measure. There are three types of validity: content, construct, and criterion. Con-

From the Southern California Permanente Medical Group, San Diego, CA. The institution of the authors has received a unrestricted research grant from Depuy/J&J, Smith Nephew, and Zimmer.

Correspondence to: Elizabeth W. Paxton, MA, Kaiser Permanente, Department of Orthopedic Surgery, 250 Travelodge Drive, El Cajon, CA 92020. Phone: 619-441-3143; Fax: 619-441-3136; E-mail: Liz.W.Paxton@kp.org. DOI: 10.1097/01.blo.0000171544.38095.77

tent-related validity refers the comprehensiveness of instrument material: how well it covers the range of functions and symptoms that may be affected by a disorder or its treatment. Construct-related validity is the ability of a tool to measure a theoretical construct or trait. Pain, function, and mental ability are examples of types of constructs. Construct-related validity is often evaluated by correlating scores of instruments designed to measure the same construct. For example, we would expect a pain questionnaire to have a high correlation with the SF-36 Bodily Pain scale and a lower correlation with the SF-36 Mental Health scale because pain and mental health are different constructs. Criterion-related validity is the ability of an instrument to predict differences in an established criterion or standard of measurement. For example, stress radiography or ligament arthrometry measures of knee laxity can be used as criteria for evaluating tools designed to evaluate knee stability.

Responsiveness is often part of the construct-validation process and is another important instrument characteristic. Responsiveness is the sensitivity of an instrument to detect clinically meaningful change. Typically, responsiveness is evaluated by comparing scores before and after an intervention that is expected to produce an observable change in pain, function or other clinically relevant outcomes.

Reliability of an instrument is another important property for consideration in selecting an outcome tool. Reliability refers to measurement error of an instrument. Reliability is expressed in terms of reproducibility, or the ability of the instrument to produce similar scores in a variety of settings in the absence of clinically significant change. Reliability also is expressed by what is called internal consistency, which is assessed by how well individual items correlate with each other.

Practical Considerations

Practical considerations also are important in selecting an instrument. For example, respondent and test-administrator burden such as average time for completion, number of items, reading level required, cost for administration, educational level of respondent and administrator, necessity or availability of scoring software, and whether an instrument can be self-administered need to be considered in selection of an outcome measure. The setting (clinic or research, private fee-for-service or university), the funding and support available, and the objectives of the study also influence the choice of an instrument.

Equally important from a practical standpoint is whether the scores are interpretable or applicable in a clinical situation. For example, if the scoring and interpretation of an outcome tool is a complex process, results may not be immediately available during a clinic visit when the information would be most useful to the patient and clinician.

Although individual scales have not been validated extensively for the purpose of making surgical decisions, they can be useful in helping patients understand how they fit with the population of patients having and considering orthopedic surgery.

Potential Outcome Measures

There are many instruments available to evaluate outcomes of PFA. Most outcome instruments fall into one of three categories: general health scales, knee-specific scales, or disease-specific scales.

General health instruments typically assess quality of life and overall health status. The Medical Outcomes Study (MOS) SF-36²⁴ and SF-12²³ are two general health measures widely used in the orthopedic literature (Table 1). The SF-36 assesses a broad array of function and well being and frequently is used in the evaluation of TKA outcomes. The SF-36 is reported to be particularly well suited to measuring outcomes related to lower limb dysfunction, in some cases showing as much responsiveness as knee-specific scales.¹⁷ Marx (e-mail correspondence, August 2004) found the SF-36 Physical Component Summary Scale scores were as responsive as the knee-specific Modems Knee Core scores after TKA. Although the SF-36 offers normative comparisons and has been validated in the TKA population,^{6,17} the scoring of the tool is difficult without a computerized system, making implementation and interpretation difficult in a clinical setting.

The SF-12, a condensed version of the SF-36, measures general health and function and also has been used to evaluate TKA outcomes. Dunbar et al⁸ compared the SF-12, SF-36, Sickness Impact Profile, Nottingham Health Profile, Lequesne, Oxford-12, and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) scores in 3,600 TKA patients. Based on instrument utility and psychometric properties, they recommended the SF-12 for large TKA databases in a cross-sectional population.

Although the SF-12 is quicker to administer and has strong reliability and validity in large samples, SF-12 scores are less precise in small samples and provide less information than the SF-36.²³ Although efficiency in an office setting would favor the SF-12, its lack of precision makes it less useful in small samples or for making treatment decisions about an individual patient. On the other hand, in very large study populations such as registries, the lower precision of the SF-12 compared with the SF-36 would be less of a disadvantage, and it offers the benefit of reducing overall respondent burden and cost of administration.

Knee scales, designed specifically to evaluate knee function, also are widely available.^{3,9,10,12,22,25} Although there are many knee scales available, few have been vali-

TABLE 1. Instruments for Evaluating PFA Outcomes

Instrument	Type of Instrument	Number of Items	Time to Complete (minutes)	Domains	Self Administered	Web Page
SF-36	General health	36	5–10	Physical function Bodily pain Role, physical General health Vitality Social functioning Role, emotional Mental health	Yes	http://www.sf-36.org
SF-12	General health	12	2–4	Physical health Mental health	Yes	http://www.sf-36.org
KSS	Knee specific	14	Approximately 20	Function Pain	No	http://www.kneesociety.org
WOMAC	Disease specific	24	4–8	Stiffness Pain Activity limitation	Yes	http://www.womac.org
KOOS	Knee specific	42	10	Pain Symptoms Activities of daily living Sport and recreation Knee related quality of life	Yes	http://www.koos.nu

SF-36 = Short Form-36; SF-12 = Short Form 12; KSS = Knee Society Clinical Rating System; WOMAC = Western Ontario and McMaster University Osteoarthritis Index; KOOS = Knee Injury and Arthritis Outcome Scale

dated for PFA and TKA outcome studies. The Knee Society Clinical Rating System¹⁰(KSS) is one instrument that has been used to evaluate both PFA¹⁴ and TKA outcomes.^{13,16,17} The KSS consists of a physical examination and functional assessment with separate knee and function scores. The KSS total score consists of a function score based on walking and stairs function and deducted points for walking aides and a knee score based on pain, range of motion, anteroposterior stability, mediolateral stability, flexion contracture, extension lag, and alignment (deductions). Although Kreibich et al¹⁶ reported the KSS scores were responsive to TKA outcomes, Lingard et al¹⁷ reported that the KSS function score was not responsive to TKA outcomes. They also indicated that the KSS total score was less responsive than the SF-36 and the WOMAC scores because it had lower correlations with patient-reported satisfaction and improvement in quality of life. Konig et al¹³ reported that the two KSS scores were responsive to TKA outcomes but differed with time. Although the knee score remained constant from 2 years on, the function score peaked at 2 years and then declined thereafter, suggesting that the scores should not be combined to create a total score. Although the KSS is widely accepted by surgeons, the instrument cannot be self administered and the reliability and validity of this tool is questionable for PFA and TKA outcome assessment.

Two well-validated knee-specific scales deserve brief discussion. The International Knee Documentation Com-

mittee (IKDC) standard form was originally developed for studying ligament deficiency.⁹ It recently has been updated to include a complete scoring system for subjective outcomes.¹¹ In a study at our institution,¹⁹ the original version had strong reliability but poor validity in evaluation of patients after acute patellar dislocation. However, the subjective portion of the newer version of the IKDC form recently was shown to have good reliability and validity in assessing subjective knee-specific outcomes in patients with a variety of patellofemoral disorders.¹¹ In the study by Irrgang,¹¹ patients with advanced patellofemoral arthritis and those having PFA or TKA were not tested specifically.

The Knee Outcome Survey (KOS)^{12,18} is reported to be useful in the evaluation of isolated patellofemoral conditions. No study has yet been published showing its value compared with the Knee Injury and Arthritis Outcome Scale (KOOS), WOMAC, or SF-36 in treating patients having TKA or PFA. The KOS and the new version of the IKDC have potential to be useful in studying patients having PFA. Further studies are needed, however, before we would recommend them in this setting because of the availability of alternatives such as the WOMAC, KOOS, and SF-36.

In addition to the knee-specific measures, there also are many disease-specific tools developed for evaluation of lower extremity osteoarthritis. The WOMAC⁴ is the most popular osteoarthritis outcome measure used in the United

States and in other countries. The WOMAC assesses lower extremity arthritis and has shown reliability and validity in the total joint replacement population⁴ (Table 1). Bombardier et al⁶ compared the WOMAC and the SF-36 in a study of 1,750 TKA patients. They reported that the WOMAC discriminated better among patients with knee problems while the SF-36 discriminated better among subjects with different levels of health and comorbidities. Based on this finding, Bombardier et al⁶ recommended using both general-health and disease-specific instruments to evaluate patient outcomes.

Although the WOMAC is used widely and has strong psychometric properties, the measure was developed for older patients with advanced osteoarthritis. Roos et al²⁰ improved on this limitation by extending the WOMAC with items pertaining to younger, active patients. Using existing WOMAC items, a literature review, content expert opinion, and a pilot study, they developed the KOOS.²⁰ The KOOS has shown good reproducibility and internal consistency, as well as strong validity and responsiveness.²¹ Another benefit of the KOOS is that it is self administered, Table 1. Considering the psychometric properties of this measure and the low respondent and administrator burden, this measure is a strong candidate for evaluating PFA outcomes.

DISCUSSION AND RECOMMENDATIONS

A variety of instruments, some with unknown reliability, validity, and responsiveness, have been used to evaluate PFA outcomes.^{1,5,7,10,14,15} The diversity of these instruments and lack of evidence showing adequate psychometric properties complicates the interpretation and generalization of PFA study outcomes. It would be useful to identify an instrument with strong psychometric properties that could be used by all clinicians and investigators in order to standardize PFA outcome assessment.

Although validity and reliability studies are lacking in evaluation of PFA outcomes, numerous studies have examined general health, knee-specific, and disease-specific instruments in evaluating TKA outcomes.^{6,8,16,17,21} The SF-36 is one of the most widely used general health measures with proven reliability and validity. It is widely available and has been validated in many languages, which would support multinational clinical collaboration. Furthermore, age-matched and gender-matched population normative data are available. Although the condensed version, the SF-12, is more efficient than the SF-36, it is less reliable than the SF-36 in smaller samples typically used in PFA outcome studies. The SF-36 shows better responsiveness even than some knee-specific scales (such as the KSS and Modems Knee Core) in evaluating TKA, as noted above.¹⁷ For practical reasons, we see no advantage in

using “just any” knee-specific scale in addition to the SF-36; if we are to invest the time in collecting and analyzing data, there should be a proven benefit.

The KSS has been used to evaluate TKA and PFA outcomes; however, the reliability and validity of this tool is questionable and the KSS cannot be self-administered. In contrast, the WOMAC, a disease-specific tool, has strong psychometric properties and can be self-administered. The KOOS improves on the WOMAC by adding items for active younger patients, who generally represent the younger PFA population.

General health, disease-specific, and knee-specific instruments measure different constructs. A general health instrument is necessary to evaluate overall health and patient comorbidities. However, because general health measures are not always highly responsive to changes in knee function, a knee-specific instrument is required to distinguish among the patients in the study population when various interventions or stratification variables are being studied.

Based on our review of the literature, we recommend the SF-36 and the KOOS for evaluation of PFA outcomes. Although both instruments are ideal for a complete evaluation, administration of both instruments during a busy clinic may not be feasible; in that case, the KOOS, which specifically is designed to evaluate knee function, is preferred.

In implementing the SF-36 and KOOS for PFA outcome assessment in the clinical setting, the first step is to request permission to use the instrument from the authors and obtain a manual containing information on psychometric properties, scoring instructions, and guidelines for administration. Typically, this is accomplished online or by directly contacting the authors. Table 1 contains URLs and contact information for the recommended PFA outcome instruments.

After obtaining the permission to use the instrument, the next step is to consider practical issues for administering the tool in your specific clinical setting. First, the cost of implementation must be considered. This will be contingent on the method of administration (self versus administrator, paper versus computerized), the staff required to administer the tool (receptionist, medical assistant, nurse, physician), and the sample size of the population of interest. The cost of implementation will often determine the data collection procedure. Piloting the instruments on a small number of patients prior to implementation can provide valuable information, assist in finalizing the data collection process and identify potential problems. Piloting the instruments also can assist in the development of a policy and procedure (P&P) or manual of operations (MOP), which provide specific guidelines for data collection. The P&P or MOP should identify the patient sample;

the individuals responsible for administering the instrument, processing the forms, entering data; how the instrument will be processed, stored, scored; the timing of data collection. A flow chart defining the process and responsible parties can clarify the data collection procedure. In addition, guidelines for addressing specific problems and answering frequently asked patient questions (eg, when a patient asks for assistance in completing the questionnaire, does not complete specific items, refuses to complete the instrument, or asks how their information will be used) are very beneficial. Once the data collection process is finalized, we strongly advise you to provide an in-service for staff involved in measuring the outcomes. It is essential that staff administer the measures according to the manual in order to obtain accurate and reliable results. We have found that in-services need to be conducted on an ongoing basis to train new personnel and adjust for any changes in the clinic setting.

In summary, identifying ideal instruments and processes for routine data collection in the clinic setting requires planning and preparation that will ultimately benefit clinicians and researchers by providing accurate and reliable information on PFA patient outcomes. Integration of PFA outcome assessment into routine care provides the highest probability of success and containment of costs. This article provides recommendations on using the SF-36 and KOOS for evaluating PFA outcomes based on a review of the literature. Further research is necessary to assess the validity and reliability and practical issues of these tools specifically relating to evaluation of PFA outcomes.

References

- Arciero R, Toomey H: Patellofemoral arthroplasty: A three- to nine-year followup study. *Clin Orthop* 236:60–71, 1988.
- Arnbjornsson A, Ryd L: The use of isolated patellar prostheses in Sweden 1977–1986. *Int Orthop* 22:141–144, 1998.
- Barber S, Noyes F, Mangine R, McCloskey J, Hartman W: Quantitative assessment of functional limitations in normal and anterior cruciate ligament-deficient knees. *Clin Orthop* 255:204–214, 1990.
- Bellamy N, Buchanan W, Goldsmith C, Campbell J, Stitt L: Validation study of WOMAC: A health instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 15:1833–1840, 1988.
- Blazina M, Fox J, Del Pizzo W, Broukhim B, Ivey F: Patellofemoral replacement. *Clin Orthop* 144:98–102, 1979.
- Bombardier C, Melfi C, Paul J, et al: Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. *Med Care* 33:AS131–AS144, 1995.
- Cartier P, Sanouiller J, Grelsamer R: Patellofemoral Arthroplasty: 2–12 year follow-up study. *J Arthroplasty* 5:49–55, 1990.
- Dunbar M, Robertsson O, Ryd L, Lidgren L: Appropriate questionnaires for knee arthroplasty: Results of a survey of 3600 patients from the Swedish Knee arthroplasty Registry. *J Bone Joint Surg* 83B:339–344, 2001.
- Hefti F, Muller W, Jakob R, Staubli H: Evaluation of knee ligament injuries with the IKDC form. *Traumatol Arthroscopy* 1:226–234, 1993.
- Insaall J, Dorr L, Scott R, Scott W: Rationale of the Knee Society clinical rating system. *Clin Orthop* 1989:13–4, 1989.
- Irrgang J, Anderson A, Boland A, et al: Development and Validation of the International Knee Documentation Committee Subjective Knee Form. *Am J Sports Med* 29:600–613, 2001.
- Irrgang J, Synder-Mackler L, Wainner R, Fu F, Harner C: Development of a patient-reported measure of function of the knee. *J Bone Joint Surg* 80A:1132–1145, 1998.
- Konig A, Scheidler M, Rader C, Eulert J: The need for a dual rating system in total knee arthroplasty. *Clin Orthop* 345:161–167, 1997.
- Kooijman H, Driessen A, Van Horn J: Long term results of patellofemoral arthroplasty: A report of 56 arthroplasties with 17 years of follow-up. *J Bone Joint Surg* 85B:836–840, 2003.
- Krajca-Radcliffe J, Coker T: Patellofemoral Arthroplasty: A 2 to 18 year follow-up study. *Clin Orthop* 330:143–151, 1996.
- Kreibich D, Vaz M, Bourne R, et al: What is the best way of assessing outcome after total knee replacement? *Clin Orthop* 331:221–225, 1996.
- Lingard E, Katz J, Wright R, Wright E, Sledge C: Validity and responsiveness of the Knee Society Clinical Rating System in comparison with the SF-36 and WOMAC. *J Bone Joint Surg* 83A:1856–1864, 2001.
- Marx R, Jones E, Answorth A, et al: Reliability, validity, and responsiveness of four knee outcome scales for athletic patients. *J Bone Joint Surg* 83A:1459–1469, 2001.
- Paxton E, Fithian D, Stone M, Silva P: The reliability and validity of knee-specific and general health instruments in assessing acute patellar dislocation outcomes. *Am J Sports Med* 31:487–492, 2003.
- Roos E, Roos H, Lohmander L, Ekdahl C, Beynon B: Knee Injury and Osteoarthritis Outcome Score (KOOS)—Development of a self-administered outcome measure. *J Orthop Sports Phys Ther* 78:88–96, 1998.
- Roos E, Toksvig-Larsen S: Knee injury and osteoarthritis Outcome Score (KOOS)—validation and comparison to the WOMAC in total knee replacement. *Health Qual Life Outcomes* 1–17, 2003.
- Tegner Y, Lysholm J: Rating systems in the evaluation of knee ligament injuries. *Clin Orthop* 198:43–49, 1985.
- Ware J, Kosinski M, Keller S: A 12-item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 34:220–230, 1996.
- Ware J, Sherbourne C: The MOS 36-item Short-Form Health Survey (SF-36). Conceptual Framework and Item Selection. *Med Care* 30:473–483, 1992.
- Windsor R, Insaall J, Warren R: The Hospital for Special Surgery knee ligament rating form. *Am J Knee Surg* 1:140–145, 1988.