ISAKOS Scientific Committee Report

Knee Rating Scales

Robert G. Marx, M.D., M.Sc., F.R.C.S.C.

Abstract: In the past 2 decades, outcome assessment following knee surgery has focused increasingly on the patient’s perspective. While traditional measures of outcome, including physical examination, imaging studies, and measures of knee laxity are complementary, questionnaires have become more important in determining the value of a procedure. Reliability, validity, and responsiveness are all important measurement qualities for health-related quality of life instruments. There are several questionnaires available, both for active patients with disorders of the knee, as well as for older patients with degenerative conditions. Activity level is also an important prognostic variable for patients with disorders of the knee. Clinical researchers should also use a validated activity rating scale to evaluate what patients are doing, in addition to how they are doing. Key Words: Knee—Quality of life—Measurement of outcome—Knee rating scales.

Traditionally, measures of success following knee surgery have been based on physical examination and radiographic variables. In the past 2 decades, outcome assessment following orthopaedic surgery has focused increasingly on the patient’s perspective. While this evolution towards the incorporation of patient-based measures is much needed, traditional measures of outcome, including physical examination, imaging studies, and measures of knee laxity, are complementary and should not be viewed as unnecessary.

Knee surgery is generally performed for symptoms and disability, and knee pain is the most common reason for which surgery is performed. Disability varies among patients who undergo knee surgery and depends to a large extent on the individual. Disability for an elite athlete may involve inability to perform at their desired level of competition. For an elderly individual with knee arthrosis, disability may involve difficulties with activities of daily living or walking.

The goal of the treatment is important to take into account when one is choosing a measure with which to evaluate the results of treatment for a disorder of the knee. If an inappropriate outcome is selected to evaluate the result of anterior cruciate ligament (ACL) reconstruction or total knee arthroplasty (TKA), incorrect treatment decisions may be made for future patients. Therefore, it is critical to use measures of clinical outcome that are important to the patients being evaluated, while also being relevant to the surgeon.

This article discusses measures of clinical outcome that may be used to evaluate different treatments for patients with disorders of the knee. The measurement properties of reliability, validity, and responsiveness are also reviewed for the reader to better select an instrument. Instruments for athletic patients with disorders of the knee are reviewed followed by instruments for patients with degenerative disorders of the knee. Lastly, measures of activity level are discussed.

RELIABILITY, VALIDITY, AND RESPONSIVENESS

A measure of any kind is only useful if it is reproducible (reliable) and accurate (valid). In the assess-
ment of health status, measures must also be able to detect improvement or worsening (termed responsiveness or sensitivity to change).

Reliability

An instrument is reliable if it measures something in a reproducible fashion. Reliability is also known as reproducibility, because repeated administrations of the same questionnaire to stable patients should produce more or less the same results. There are two schools of thought with respect to the measurement of reliability for health-status instruments. The first is test-retest reliability, which involves having patients who are in a stable state respond to the questionnaire at 2 points in time. The time period must not be too short because the subject will remember their prior responses. As well, the time period must not be too long, which would allow for the possibility of clinical change. In general, a time period ranging from 2 days to 2 weeks is used.

Measures of agreement such as the intraclass correlation coefficient and/or the limits of agreement statistic are typically used to compare the scores. The intraclass correlation coefficient is an index of concordance for dimensional measurements ranging between zero and one, where ≥0.75 is adequate for patients enrolled in a clinical trial. This statistic is important to differentiate from measures of correlation, such as the Spearman or Pearson correlation coefficients, which do not measure agreement. These statistics may indicate excellent correlation in situations where agreement is poor and, therefore, they should not be used for studies of reliability. For example, if the first measure is twice as high as the second measure for all subjects in a study of reliability, the correlation would be perfect but the agreement would be poor.

The limits of agreement statistic is a descriptive measure of reproducibility. This value is the mean difference between the 2 tests ±2 standard deviations. Ninety-five percent of the differences between the 2 test administrations will lie within this interval, providing the investigator with an estimate of the precision of the measure.

Internal consistency is another method for measuring the reliability of rating scales. This concept was borrowed by clinicians from the field of psychometrics. The latter discipline involves the measurement of psychological phenomena (such as depression or anxiety) or educational achievement.

The concepts evaluated by psychometric scales are difficult to define and/or may involve learning. In these situations, it would not be possible to have the patients complete the questionnaire on 2 separate occasions, due to recall or learning effects. The calculation of internal consistency involves a measurement of the intercorrelation of the responses to the questions on a single administration. The statistic generally used to describe internal consistency is termed Cronbach's alpha, which ranges from zero to one, with one indicating perfect reliability. Cronbach's alpha has been used to evaluate the reliability of rating scales in orthopaedic surgery. However, it is questionable whether the principles of Psychometric theory apply to the measurement of symptoms and disability. In practice, orthopaedic scales that measure a wide variety of clinical phenomena have also been demonstrated to have high internal consistency.

Validity

An instrument is valid if it measures what it is intended to measure. There are several types of validity, which are reviewed briefly below.

The simplest way of validating a rating scale is to provide evidence that its results match a gold standard. This is known as criterion validity, although it is generally not possible for instruments that measure quality of life. In such situations, we must rely on face validity, content validity, and construct validity.

Face validity is present when an expert clinician reviews the questions in the scale and believes that they appear to measure the concept in question. This form of validity is rather simple, and nevertheless important.

Content validity is a more formal application of face validity. Content validity measures whether the scale includes representative samples of the concept that the investigator is attempting to measure. For example, if a rating scale were measuring quality of life, the content of the scale should include measures of physical, mental, and social health to provide adequate content validity.

Construct validity determines whether the questionnaire behaves in relation to other measures as would be expected. This requires several hypotheses about how the results of the questionnaires should correlate (positively or negatively) with other related or unrelated measures, and in testing these hypotheses.

Responsiveness

Orthopaedic surgeons generally use rating scales to measure improvement in health-related quality of life
Knee Rating Scales

Knee Rating Scales for Athletic Patients

There are many rating scales available to measure outcome in athletic patients with disorders of the knee. What defines an athletic individual may be not always be clear. The activity level of the patient is an important prognostic variable, because active patients place greater demands on their knees than sedentary individuals and have different expectations of the results of treatment. Activity level is not always directly related to symptoms and disabilities, and should be measured separately. This topic is discussed at the conclusion of this article.

In a recent article, Weitzel and Richmond review and evaluate a large number of knee scales that are mainly used for the sports medicine population. A discussion of 8 of these commonly used rating scales for athletic patients with disorders of the knee is presented below.

The modified Lysholm scale is an 8-item questionnaire originally designed to evaluate patients after knee ligament surgery. It is scored on a 100-point scale, with 25 points attributed to knee stability, 25 to pain, 15 to locking, 10 each to swelling and stair climbing, and 5 each to limp, use of a support, and squatting. Although this scale was developed without direct patient input, it has been used extensively for clinical research studies and has adequate test-retest reliability and good construct validity.

The first version of the Cincinnati Knee Rating System was published in 1983, with additional modifications that were developed for occupational activities, athletic activities, symptoms, and functional limitations with sports and daily activities. There are 11 components in the Cincinnati Knee Rating System. In addition to measuring symptoms and disability, there are sections of this rating system that measure physical examination, laxity of the knee based on instrumented testing, and radiographic evidence of degenerative joint disease. This instrument is reliable, valid, and responsive to clinical change.

The American Academy of Orthopaedic Surgeons (AAOS) Sports Knee Rating Scale was included in the Musculoskeletal Outcomes Data Evaluation and Management System (MODEMS) for athletic patients with disorders of the knee. There are 5 parts and 23 questions in this instrument: a core section including stiffness, swelling, pain, and function (7 questions), a locking or catching on activity section (4 questions), a giving way on activity section (4 questions), a current activity limitations due to the knee section (4 questions), and a pain on activity due to the knee section (4 questions).

The 5 subscales are independent and are meant to be reported separately. As well, this scale has the response "cannot do for other reasons" for many questions. The scoring manual states that an item should be "dropped" if the patient selects that response, which may be interpreted as "scored as missing." These factors may lead to practical difficulties when using this questionnaire. Despite these concerns, the measurement properties of this instrument were found to be satisfactory when the 5 subscales were combined and the mean was calculated.

The Activities of Daily Living Scale of the Knee Outcome Survey was published with an evaluation of its reliability, validity, and responsiveness. It was developed based on a review of relevant instruments and clinician input. This scale is designed for patients with disorders of the knee ranging from ACL injury to arthrosis. It includes 17 multiple choice questions divided into 2 sections, 1 for symptoms (seven questions) and 1 for functional disability (10 questions). This instrument has acceptably high correlation with the Lysholm, Cincinnati, and AAOS scales, as well as other measures of disability, indicating excellent construct validity. It was also found to be slightly more sensitive to clinical improvement (responsive) than the three other scales in a group of athletic patients.

The single assessment numeric evaluation (SANE) was devised to evaluate college-age patients following ACL reconstruction. The SANE asks the patient how they would rate their knee, from zero to 100, with 100 being normal. This score was found to correlate well with the Lysholm scale in this patient population. The advantage of this single question is its simplicity and the ease with which it can be administered. One potential pitfall is that a single, relatively broad question may be interpreted differently by patients with different disorders and varying levels of
symptoms and disability. In the setting of a very homogeneous cohort, such as college-age patients recovering from a specific procedure (such as ACL reconstruction), the range of pathology is relatively narrow and the instrument correlates well with a standard measure of knee function. The applicability of this tool to preoperative patients or patients with a variety of diagnoses is unknown.

The Knee Injury and Osteoarthritis Outcome Score (KOOS) was developed with input from patients who underwent remote meniscal surgery.24 The reliability, validity, and responsiveness was determined to be satisfactory in a cohort of 21 patients who underwent ACL reconstruction.24 Five separate scores are calculated for pain, symptoms, activities of daily living, sport and recreation function, and knee-related quality of life. Of particular interest, the WOMAC, discussed in greater detail below, is included in the KOOS and its score can be determined from the KOOS.

The quality of life outcome measure for chronic ACL deficiency was developed by Mohiadi.35 This instrument was developed by surveying ACL-deficient patients, primary care sports medicine physicians, orthopaedic surgeons, athletic therapists, and physical therapists. The scale comprises 31 visual analog questions regarding symptoms and physical complaints, work-related concerns, recreational activities and sport participation, lifestyle, and social and emotional health status relating to the knee. This rating scale is valid and responsive for patients with ACL insufficiency.35 It is very specific to ACL deficiency and thus would not be applicable to other disorders of the knee.

The International Knee Documentation Committee (IKDC) developed a rating scale for 7 “objective” parameters relating to the knee.36 These included effusion, motion, ligament laxity, crepitus, harvest-site pathology, radiographic findings, and 1-leg hop test. Patients were graded as normal, nearly normal, abnormal, or severely abnormal on each of these. The lowest grade for a given group determines the final patient grade.

More recently, the IKDC has developed a questionnaire relating to “subjective” factors. These include symptoms, sports activities and ability to function, including stairs, squatting, running, and jumping. It is currently available on the Web site of the American Orthopaedic Society for Sports Medicine at www.sportsmed.org/Research/Default.htm. At the time of this writing, the reliability, validity and responsiveness testing has been completed, and publication is pending.37

KNEE RATING SCALES FOR PATIENTS WITH DEGENERATIVE DISORDERS OF THE KNEE

There are several knee rating scales available for patients with arthrosis of the knee. These rating scales were generally designed to evaluate patients with a greater level of disability than the scales reviewed in the previous section. The 3 scales discussed below are commonly used to evaluate patients following TKA.

The most commonly used instrument for patients with knee arthrosis is the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC).38-40 This scale involves 24 questions: 5 relating to pain, 2 relating to stiffness, and 17 relating to difficulty with activities of daily living. This scale is responsive and valid for patients with arthrosis.41-43 The WOMAC has been translated into many languages and these versions have been validated as well.40,44,45

The index of severity for knee disease46 was initially developed for nonsteroidal anti-inflammatory drug trials. This questionnaire involves 5 questions related to pain, 1 question related to the maximum distance the patient can walk, and 4 questions relating to activities of daily living. This scale was initially intended to be interviewer-administered, although a questionnaire format has subsequently been validated as well.43

Part of the MODEMS package includes a knee core rating scale. This section, which includes 7 questions, is recommended for use in patients with osteoarthritis of the knee. The questions relate to knee stiffness, knee swelling, use of a support to get around, putting on socks, and pain with walking, climbing stairs, and lying in bed at night. This core scale is included in the AAO Sports Knee Rating Scale as one of the 5 subscores. There are no published data supporting the reliability, validity, or responsiveness of this scale for use in patients with osteoarthritis of the knee.

The Oxford Knee Scale was developed using patient input to select the most relevant items.47 The developers of this tool interviewed multiple groups of 20 patients who were attending an outpatient clinic for consideration of TKA to determine which questions should be included. After each group of 20 patients, the investigators modified the responses and retested the items. The questionnaire is comprised of 12 multiple-choice questions, each with 5 responses. It was tested in a prospective group of 117 patients undergoing TKA and is reliable, valid, and responsive.47,44
MEASURES OF ACTIVITY LEVEL FOR PATIENTS WITH DISORDERS OF THE KNEE

Patients' activity levels are related to prognosis in the sports medicine population since people who are very active have different expectations and demands than those who are relatively sedentary. Because the frequency and intensity of sports participation varies widely among patients, a measure of activity is important for studies evaluating such individuals. For example, a study describing a new surgical technique for a knee disorder should document the patients' activity level to ensure that the results can be applied to the appropriate patient population. For studies comparing 2 groups of patients, it is important for the activity levels of the 2 groups to be similar to avoid a biased comparison.

In a systematic literature review, S activity level rating scales that are potentially applicable to outcome studies in sports medicine were identified. There were inherent problems with each of the available instruments, which led to the construction of a new rating scale for this purpose. This activity rating scale consists of 4 questions relating to the frequency with which the patient runs, cuts, pivots, and decelerates. It has been shown to be reliable and valid. This scale is recommended in addition to a knee outcome instrument for the evaluation of athletic patients with disorders of the knee.

CONCLUSIONS

There are several reliable, valid, and responsive instruments that can be used to evaluate the outcome of surgery for disorders of the knee in both athletic patients and individuals with degenerative conditions. It is also important to document their level of activity with a valid instrument, as this is an important prognostic variable for these patients.

REFERENCES


