Level V Evidence

Measuring Arthroscopic Outcome

James J. Irrgang, M.D., and James H. Lubowitz, M.D.

Abstract: In order to measure and report clinical research results, we require the ability to measure our results, and we label our measurement tools “outcome measures.” Options are many and include clinician-reported and patient-reported measures. Patient-reported measures include measures of symptoms (such as pain), measures of activity and function (such as work ability or sports participation), and measures of general health status (such as quality of life). Measures of symptoms and measures of activity and function may be specified as joint or region, and disease or injury, specific. We recommend that clinical outcome studies include a combination of these measures. In addition, clinical outcome measures should be practical, widely accepted, reliable, valid and responsive. Finally, the evidence for reliability, validity and responsiveness should be specific to the disease and/or population of interest. Key Words: Outcomes research—Outcome measure—Patient reported—Clinician reported—Valid—Health status.

Whether consenting a patient for a procedure an arthroscopist has performed 1,000 times, or telling a colleague about an intervention an arthroscopic surgeon has performed only once, the question will be the same: “What are your results?” To report our results, we require the ability to measure our results, and we label our measurement tools “outcome measures.”

Next, we need to describe our results to patients or colleagues. Thus readers of Arthroscopy and particularly contributors of original scientific articles must understand how outcomes are best measured and reported in the orthopaedic, arthroscopic, and medical literature. The options are many and include the following:

- Measures recorded by the clinician that are joint-specific such as range of motion,
- patient-reported measures of symptoms such as use of a visual analog scale to measure pain,
- patient-reported measures of function,
- patient-reported measures of general health status such as quality of life.

Patient-reported measures are sometimes called “subjective” outcome measures, and clinician-measured outcomes are sometimes called “objective” outcomes; however, the term “subjective” may have a negative connotation (if misinterpreted as an antonym to “objective”) and is thus best avoided (Robert Marx, M.D., personal communication, October 2006). In fact, the patient-reported measures may indeed be considered more objective than a clinician-derived measurement because most have well-documented reliability, validity, and responsiveness. Rather, the term “patient-reported outcome” is preferred to the term “subjective outcome measure.”

It is of utmost importance to understand that the relation between impairment of the joint as measured by the clinician and patient-reported outcome is not direct. For example, several authors have shown that after anterior cruciate ligament injury and reconstruction, there is no relation between anterior laxity measured with a knee arthrometer and patient-reported outcome.1,2 Thus, in the scientific community, there is...
an emerging consensus that a patient’s perspective regarding the outcome of a surgical or other intervention is of primary concern.\textsuperscript{3-6}

In addition, a consensus is emerging in the fields of arthroscopy and orthopaedic sports medicine that measures of return to activity (including work and sports participation) are of utmost importance.\textsuperscript{7} Because it is difficult to reliably and accurately observe and rate an individual’s performance of activities, we again note that an appreciation of the importance of patient-reported measures of activity continues to gain in significance.

\section*{PATIENT-REPORTED OUTCOME MEASURES}

Clinician-reported outcome measures (physical examination or imaging study findings) are well known to readers of \textit{Arthroscopy}. Therefore, we focus this section on patient-reported outcome measures.

Patient-reported outcome measures can be classified as general or specific measures of health status. General health status outcome measures are designed to be applicable across a number of disease processes and interventions and across demographic and cultural subgroups. These measures are designed to give a comprehensive overview of the patient’s health by measuring multiple dimensions of physical and mental health. The most widely known and accepted general health status outcome measure is the Medical Outcomes Study Short Form (SF) 36.\textsuperscript{8-10} A more concise 12-question version of this form (SF-12) takes less time for a patient to complete. Because it includes relatively few questions, the SF-12 is less useful for making decisions about a specific patient (e.g., Is an individual patient better or worse?), but it is acceptable for measuring the outcome in terms of physical and mental health of a group of patients.\textsuperscript{11}

Although general outcome measures of health status permit comparison across populations with different health conditions, an important limitation of general outcome measures of health status for readers of and contributors to the arthroscopic literature is that general outcome measures tend to be less responsive to change in the patient’s status than specific patient-reported outcome measures. In addition, by definition, general measures of health status evaluate a broad range of health including physical and emotional health. Thus the content of these measures may be less relevant to patients undergoing arthroscopic surgery or to clinicians performing it.

Disease- or joint-specific (region-specific) patient-reported outcome measures are designed to focus on aspects of health that are directly related to the primary condition or population of interest with the intent of creating a more responsive outcome measure; when the disease improves, a significant change in the measure should reflect the change in health. Disease-specific outcome measures are developed for a particular injury or illness. The content of these outcome measures includes symptoms, activity limitations, and participation restrictions commonly experienced by individuals with the particular injury or illness for which the instrument was designed. An example of a disease-specific outcome measure is the Western Ontario and McMaster’s University Osteoarthritis Index (WOMAC)\textsuperscript{12} for patients with hip or knee osteoarthritis.

Joint- or region-specific patient-reported outcome measures have been developed to determine the effects of a variety of injuries affecting a particular joint or body area. The content of region-specific patient-reported outcome measures reflects the symptoms, activity limitations, and participation restrictions commonly experienced by individuals with injury or impairment of the particular region for which the instrument was developed. Examples of region-specific patient-reported health status outcome measures are listed in Table 1. The advantage of using a region-specific measure as opposed to a disease-specific measure is that region-specific measures are appropriate for a wide variety of injuries or impairments affecting a particular region. This is important because patients may have more than 1 underlying injury affecting the region. For example, the International Knee Documentation Committee Subjective Knee Form is a valuable outcome measure for a patient with an anterior cruciate ligament–deficient knee with (or without)

\begin{table}[h]
\centering
\caption{Examples of Region-Specific Patient-Reported Health Status Outcome Measures}
\begin{tabular}{ll}
\hline
Region & Instrument \\
\hline
Upper extremity & Disabilities of the Arm, Shoulder and Hand (DASH) index\textsuperscript{15} \\
Shoulder & American Shoulder and Elbow Surgeons Patient Self-Evaluation Form\textsuperscript{16} \\
& Simple Shoulder Test\textsuperscript{17} \\
Low back & Oswestry Low Back Pain Disability Questionnaire\textsuperscript{18} \\
& Quebec Back Pain Disability Scale\textsuperscript{19} \\
Lower extremity & Lower Extremity Function Scale\textsuperscript{20} \\
Knee & Knee Outcome Survey\textsuperscript{21} \\
& International Knee Documentation Committee Subjective Knee Form\textsuperscript{13,14} \\
Foot/ankle & Foot and Ankle Ability Measure\textsuperscript{22} \\
\hline
\end{tabular}
\end{table}
coexisting meniscus pathology and degenerative arthritis.

Disease- and region-specific patient-reported outcome measures are responsive to small changes in a patient’s condition and are easy to administer and interpret. The increased responsiveness of specific patient-reported outcome measures results from the inclusion of questions that reflect only those symptoms, activity limitations, and participation restrictions that are relevant to the condition or region being studied. Although specific patient-reported outcome measures do not measure all aspects of health status and do not allow for comparisons between different disease states or populations (or both), specific measures have the advantage of being most likely to be relevant to patients and clinicians.

**SELECTION OF CLINICAL OUTCOME MEASURES**

Multiple factors need to be considered when selecting a clinical outcome measure. Among these factors are basic or practical considerations, psychometric considerations (such as reliability, validity, and responsiveness), and consideration of the purpose for which the information will be used. In fields such as arthroscopy, our purpose is often to quantify changes that occur in patients (with specific problems) over time. As such, it is important to select an outcome measure that has shown reliability, validity, and responsiveness for the population of patients or disease state (or both) for which the outcome measure will be used.

Practical considerations for selection of a clinical outcome measure include ease of use, acceptability by the clinical and scientific community, and cost. Patient-reported outcome measures are attractive because they require minimal administrative resources. The time required for administration and scoring and the effort required to interpret and use the data are other considerations that affect ease of use. The application of a widely accepted outcome measure ensures the ability to compare the results with other published studies, which is required to minimize reporting bias. Excessive cost for administering and scoring a clinical outcome measure will discourage its use.

Psychometric properties of an outcome measure include reliability, validity, and responsiveness. Reliability is related to the consistency and precision of measurement. Validity is related to the accuracy of the measurement. Responsiveness is related to the ability of the outcome measure to detect change. More generally, validity is defined as any evidence that lends support for interpretation of an outcome measure. As such, reliability and responsiveness of an outcome measure that is designed to measure change over time provide evidence for validity of that measure.

Test-retest reliability is the degree to which the outcome score remains stable when there is no change in the condition of the patient or population. The degree of test-retest reliability is estimated with either the intraclass correlation coefficient for continuous outcome measures or the Cohen κ coefficient for discrete outcome measures. The intraclass correlation coefficient can be used to calculate the standard error of a measurement, an indication of precision. Generally, we prefer that patient-reported outcome instruments have test-retest reliability coefficients greater than 0.8.

Responsiveness represents the ability of a measure to accurately and reliably evaluate change over a period of time. Factors affecting the magnitude of change include the population studied, the type of treatment, and the timing of data collection. For example, more rapid change would be expected in individuals recovering from surgery than in individuals with a chronic, untreated condition. Therefore responsiveness must be considered in the context of the circumstances in which the outcome measure is used. Statistics that summarize responsiveness include average change, effect size, and standardized response means. Effect size and standardized response means relate average change either to variability of initial measurements or to change measurements. Generally, we prefer instruments with a large effect size (>0.8).

Readers and reviewers of individual studies should consider the distinction between statistical significance and clinical significance or clinical relevance. Statistics such as effect size or standardized response mean do not allow for this type of interpretation. We thus consider minimal detectable change (MDC) and minimal clinically important difference (MCID). The MDC represents the amount of change necessary to be certain that the change is greater than the measurement error. Calculation of the MDC is based on the test-retest reliability coefficient and the SD of the scores or results. Because it is based on the test-retest reliability coefficient, the MDC is dependent on the conditions that were used to estimate test-retest reliability (such as timing of follow-up). The MDC for some common patient-reported outcome measures for the knee is reported in Table 2.
The MCID differentiates a patient who perceives himself or herself to be improved from a patient who does not. The MCID for the International Knee Documentation Committee Subjective Knee Score has been determined \(^\text{13,14}\): if the change score is less than 11.5, a patient does not perceive improvement; if the change score is greater than 20.5, the patient perceives improvement; and if the change score is between 11.5 and 20.5, the patient may or may not perceive improvement.

Validated outcome measures require that evidence supports the use of an outcome measure for a specific purpose. For an outcome instrument that is used to measure change over time, this must include evidence for reliability and responsiveness. Evidence for reliability is necessary to show that the outcome score changes when the patient’s condition has not changed, and evidence for responsiveness is necessary to show that the outcome score changes as the patient’s condition improves or gets worse. Thus readers must understand that a clinical outcome measure should not be considered “validated” in and of itself. Rather, as noted previously, a measure must be validated for a specific purpose such as for the evaluation of a population of subjects with a specific condition.

CONCLUSIONS

We prefer clinical outcome studies that include a combination of (1) clinician-measured outcomes that are joint-specific (such as range of motion); (2) patient-reported measures of joint symptoms (such as pain); (3) injury-, joint-, or region-specific patient-reported measures of activity and function (such as work ability or sports participation); and (4) patient-reported measures of general health status (such as quality of life). These measures should be practical, widely accepted and reported, reliable, valid, and responsive. The evidence for reliability, validity, and responsiveness should be specific to the disease or population of interest (or both) and the intended application of the outcome measure.

REFERENCES


