Clinical Outcome Assessment involves the use of clinician-based and patient-based measures. Athletic trainers are most familiar with clinician-based measures, because they frequently relate to impairments that are recorded for the “objective” section of SOAP-format documentation (e.g., range of motion and strength). Patient-based measures are typically self-report questionnaires that ask questions regarding the patient’s perception of his or her condition, injury, and/or overall health status. Ideally, clinician-based and patient-based outcome measures should be included in all initial and follow-up patient evaluations; however, the large number of available instruments may overwhelm clinicians who have some degree of apprehension about collecting patient-based outcome measures.

Patient-based outcome measures are classified as either generic or specific. Generic measures are suitable for a wide variety of patient populations and injuries; they are broad in scope and typically focus on health-related quality of life. In contrast, specific outcome measures focus on a particular region, disease/condition, or health dimension. Region-specific scales address the effects of injury/illness on a specific body region (e.g., shoulder), whereas disease/condition-specific measures are useful for evaluation of medical conditions (e.g., shoulder instability). Dimension-specific instruments are focused on one aspect of an individual’s health, such as pain. Optimal patient care requires evaluation of both generic and specific outcomes; however, there are far fewer generic measures than specific measures. For example, there are over 25 patient-based shoulder-specific outcome measures. It is important for the practicing clinician to determine which of these instruments is most appropriate for his or her clinical practice, or most appropriate for individual patients, which is challenging. The purpose of this report is to present eight criteria deserving consideration when selecting a patient-based outcome measure (Table 1).

In Instrument Development

The initial concern regarding any patient-based outcome measure is the procedure used for instrument development. Development of patient-based measures is complex, involving item generation and initial item reduction, field testing and final item reduction, and establishment of scale measurement properties. A clinician should determine whether a systematic process was used to develop this instrument. For example, the International Knee Documentation Committee Subjective Knee Form (IKDC Subjective Knee Form) was developed through a comprehensive process that began with the identification of the purpose of the instrument and the constructs to be measured. Questions for the initial version of the instrument, a total of 27, were developed by the committee through review of other knee instruments, with the goal of highlighting symptoms and functional capabilities related to activities of daily living and sports. Following question generation, a series of procedures were performed, including field testing of preliminary questions, determination of measurement properties of the initial questions, identification of final instrument questions, and field testing of the final instrument. As a result of the systematic development process, the IKDC Subjective Knee Form is a reliable and valid instrument for...
quantification of symptoms and function for a wide variety of people suffering from injuries and conditions affecting the knee. The IKDC is an example of a well-developed patient-based outcome measure practical for use by athletic trainers.

**Appropriateness**

Appropriateness of a scale refers to how closely the measure relates to a specific clinical question. A clinician should identify a particular purpose for the use of the scale (e.g., rehabilitation goal), which should relate to the injury, disease/condition, or dimension of interest. A clinician should determine whether the content of the instrument is appropriate for the clinical question, condition, or patient being treated, and whether the instrument is appropriate for a variety of patients seen in clinical practice. For example, the Anterior Knee Pain Scale (AKPS) is an appropriate evaluation tool for a patient with anterior knee pain related to patellofemoral pain syndrome and dislocation or subluxation of the patella; however, the AKPS is limited to questions related to pain and function of the knee. In contrast, the Lower Extremity Functional Scale (LEFS) is broader in scope, and is useful for a variety of patients and lower extremity conditions. The LEFS is appropriate for patients with hip osteoarthritis, ankle fractures, patellofemoral knee pain, and muscle strains. In contrast, the AKPS is not appropriate for people with hip osteoarthritis or ankle fractures.

Clinicians should not select an instrument on the basis of title alone, because the title cannot precisely identify the construct that is measured by an instrument. In terms of physical function, one instrument may ask questions specific to the amount of assistance a person needs to complete particular tasks, whereas another scale may ask questions regarding the degree of difficulty the person has in completing tasks. The representations of physical capabilities derived from these instruments may not be the same. A review of the instruments, beyond title alone, will assure that an instrument is appropriate for its intended use.

**Reliability**

Reliability is the degree to which an instrument measures the true scale score, as opposed to random error. Clinicians should determine whether the instrument produces results that are reproducible and internally consistent. Reproducibility or “stability” can be measured through administration of the instrument on two separate occasions, during an interval in which patients’ conditions are not expected to change. Typically, reproducibility is represented by the Interclass Correlation Coefficient (ICC), with values above 0.7 considered acceptable. Internal consistency relates to the homogeneity of questions in the same domain and their ability to measure the same construct. Internal consistency is frequently evaluated with Cronbach’s alpha, which is generally deemed acceptable at values of 0.7-0.9.

**Validity**

Validity refers to the ability of an instrument to measure what it intends to measure. Several types of validity are established for instruments, and clinicians should determine whether a particular instrument demonstrates an acceptable range of validity for a target population (Table 2). Although an instrument may be validated for a particular population or injury, validity is not directly transferable to other populations or injuries. The Disabilities of the Arm, Shoulder, and Hand (DASH) scale is well validated for patients ranging in age from 40-58 years old with upper extremity complaints, but to our knowledge, there are no studies validating its use for an adolescent athletic population.

<table>
<thead>
<tr>
<th>Table 1. Eight Criteria for Selecting Patient-Based Outcome Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Instrument development</td>
</tr>
<tr>
<td>2. Appropriateness</td>
</tr>
<tr>
<td>3. Reliability</td>
</tr>
<tr>
<td>4. Validity</td>
</tr>
<tr>
<td>5. Responsiveness</td>
</tr>
<tr>
<td>6. Interpretability</td>
</tr>
<tr>
<td>7. Acceptability</td>
</tr>
<tr>
<td>8. Feasibility</td>
</tr>
</tbody>
</table>
Responsiveness

An instrument is responsive if it is able to detect change over time when meaningful change has occurred. Thus, the instrument measures improvement when a patient’s condition improves and deterioration when a patient’s condition declines. Responsiveness must not be overlooked in scale selection, because an instrument can be valid and reliable, yet unresponsive to a meaningful change in status. Responsiveness can be measured many ways, such as with a change score, effect size (ES), standardized response mean (SRM), sensitivity and specificity to change, and receiver-operating characteristics. Effect size and the SRM are the most common and use similar formulae for their calculation. Although there is no established criterion for interpretation of ES or the SRM, it is generally accepted that the greater the magnitude of change, the more responsive the measure. A change of 0.8 is considered acceptable.

Interpretability

Interpretability refers to how easy it is for a clinician to understand the calculated score derived from an instrument. One criticism of patient-based outcome measures is that their scores are not as interpretable as other clinical measures, such as blood pressure and heart rate. An approach to increase the interpretability of these instruments is the reporting of the minimal detectable change (MDC) and/or the minimal important difference (MID). The MDC represents the error associated with administration of the instrument on more than one occasion, whereas the MID represents the smallest difference that a patient perceives as clinically meaningful. These measures indicate the amount of change necessary for the change to be greater than measurement error (i.e., MDC) or indicative of clinically meaningful change in status (i.e., MID).

Acceptability

Acceptability refers to the “patient friendliness” of the instrument, which includes patient perception of the instrument and response rate (i.e., completeness of the questionnaire responses). Clinicians should evaluate patient acceptability by ensuring that the instrument can be completed in a relatively short amount of time; the instrument contains clear, concise, and easy to understand questions; and that patients will be comfortable answering all questions. Short questionnaires that are easy to read limit patient burden, which is likely to provide a high response rate.

Feasibility

“Clinician friendliness” or administrative burden refers to the feasibility of administering the instrument as part of the routine care delivery process. Feasibility relates to ease of instrument administration, amount of time needed to train clinicians for administration, and

---

**Table 2. Types of Validity**

<table>
<thead>
<tr>
<th>Validity</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>Measure of the instrument’s comprehensiveness (i.e., how well do the questions reflect the purpose of the instrument?).</td>
</tr>
<tr>
<td>Face</td>
<td>Measure of how well the instrument appears to measure what it is intended to measure.</td>
</tr>
<tr>
<td>Construct</td>
<td>Quantitative assessment of whether a construct (e.g., pain) is related to another similar variable (e.g., pain medication use).</td>
</tr>
<tr>
<td>Convergent</td>
<td>Occurs when two instruments that measure similar concepts are highly correlated.</td>
</tr>
<tr>
<td>Divergent</td>
<td>Occurs when two instruments that, although similar, are not highly correlated when they measure different concepts.</td>
</tr>
<tr>
<td>Criterion</td>
<td>Measure of whether an instrument is highly correlated with a “gold standard” measure of the same theme.</td>
</tr>
<tr>
<td>Concurrent</td>
<td>Measure of whether an instrument can accurately predict an individual’s current status.</td>
</tr>
<tr>
<td>Predictive</td>
<td>Measure of whether an instrument can predict the future status of an individual.</td>
</tr>
</tbody>
</table>
costs or time associated with administration, scoring, and score interpretation. For example, some scales require purchase of a license before use (e.g., Short Form Health Survey Questionnaire, SF-36), whereas others are considered public domain (e.g., DASH). Feasibility is typically not reported in the literature, so the clinician must evaluate the administrative burden of the instrument when considering its use in his or her practice.

**Summary**

Patient-based outcome measures are intended to enhance patient evaluation and injury management by providing patient-centered information regarding function, disability, and overall quality of life. Clinicians should avoid indiscriminate use of patient-based outcome scales by first determining the quality of the instrument on the basis of the criteria presented. Improperly selected patient-based outcome measures may fail to yield appropriate information for treatment planning, clinical decision making, and development of rehabilitation goals. Unfortunately, there is no scale that ideally satisfies all instrument criteria. Therefore, clinicians must consider the relative advantages and limitations of various outcome measures and then select the instrument that best fulfills the purpose of outcome data collection.

**References**


Alison R. Snyder and Tamara C. Valovich McLeod are with the Department of Interdisciplinary Health Sciences at A.T. Still University in Mesa, AZ.