Beyond the Basics of Clinical Outcomes Assessment: Selecting Appropriate Patient-Rated Outcome Measures for Patient Care

Alison Snyder Valier, PhD, ATC  Kenneth C. Lam, ScD, ATC
Associate Professor  Assistant Professor

Department of Interdisciplinary Health Sciences
A.T. Still University
Mesa, AZ

© Valier and Lam, 2013

Introduction

• 2011 Educator’s Conference
  – Introduced concepts and application of AT Educational Competencies, 5th Edition
  – EBP-11-14
    • Disablement Models
    • HRQOL
    • Patient-Rated Outcome (PRO) measures

Which PRO should I use?

Essential Questions

Developed on sound theory and with a systematic process?
Reproducible and internally consistent results?
Measure what it claims to measure?
Measure changes over time that matter to patients?
Easy to interpret and apply?
Good score precision?
Practice Questions

Appropriate for my purpose?

Cost and time manageable?

Patients like it?

Clinicians use it?

Criteria for Selecting PROs

1. Instrument Development
2. Reliability
3. Validity
4. Responsiveness
5. Interpretability
6. Precision
7. Acceptability
8. Feasibility
9. Appropriateness

Instrument Essentials

Clinical Utility

Assignment Basics

Please select and obtain 2 specific PRO measures that appear relevant to your clinical case scenario for critique, comparison, and presentations to the class.

☑ Select and obtain the instruments
☑ Create and deliver a power point presentation
☑ Develop a narrative about selection process
☑ Contribute to key considerations Excel document (group activity)
Example Case Scenario

You are interested in measuring the quality of your clinical care in the collegiate setting, especially as it related to patients who have suffered knee injuries.

Select 2 specific PROs:

1. Knee Injury and Osteoarthritis Outcome Score (KOOS)
2. International Knee Documentation Committee (IKDC) Subjective Knee Form

(Smith, 2012; Collins, 2011; Wang, 2010; Wright, 2009; Tanner, 2007)

Instrument Development

Item generation and reduction → Testing and item reduction → Establishment of measurement properties

- Input from experts and patients
- Disablement model frameworks
- Tested in injured and healthy populations

*Find the validation article(s)
METHODS
Development of the IKDC Subjective Knee Form included specification of the purpose of the instrument, defining the construct (that is, the underlying unobservable concept; see Appendix 1) to be measured, generation, review, and preliminary type of question; field testing of preliminary questions on a large representative sample; determination of statistical properties of the questions; selection of questions for the final version of the form; and administration of the final version of the IKDC Subjective Knee Form to a sample of patients to provide evidence for reliability and validity. These activities were performed under the auspices of the IKDC with the advice of a psychometric consultant.

Irgang et al. 2001

Instrument Development:
Assignment

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Rigorous Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOOS</td>
<td>Y</td>
</tr>
<tr>
<td>IKDC</td>
<td>Y</td>
</tr>
</tbody>
</table>
Reliability: Relevance to PROs

Reliability Coefficients
• Reflects how well an instrument can differentiate between people
• Important for measuring change over time

Population Specific
• Not inherent characteristic of instrument
• High school vs. outpatient clinic
• High school athletes vs. middle aged individuals

Acceptable Values
• Group = .70
• Individual = .9

(Streiner and Norman 2008, Nunnally 1978)

Reproducibility

• Do scores on an instrument remain the same when a patient’s health status is unchanged?
• Common Measures
  – Pearson product moment correlation coefficient (r)
  – Intra-class correlation coefficient (ICC)
• Values
  – Individuals = .9
  – Groups = .7

(Streiner and Norman 2008, Nunnally 1978)

Internal Consistency

• More than one question used to measure a dimension of health (eg, physical, social)
  – All questions related to a health dimension should be homogenous
• Chronbach’s Alpha
  – Higher alpha=higher internal consistency
  – Acceptable range = .7-.9
• Too perfectly correlated produces narrow measure of a dimension

© Valier and Lam, 2013
**Internal Consistency: IKDC Example**

9. How does your knee affect your ability to

<table>
<thead>
<tr>
<th></th>
<th>Not difficult at all</th>
<th>Minimally difficult</th>
<th>Moderately difficult</th>
<th>Extremely difficult</th>
<th>Unable to do</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Go up stairs</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>b. Go down stairs</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>c. Kneel on the front of your knee</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>d. Squat</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>e. Sit with your knee bent</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>f. Rise from a chair</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>g. Run straight ahead</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>h. Jump and land on your involved leg</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>i. Stop and start quickly</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

Better internal consistency for function means that the function questions score similarly.

Knee injury and Osteoarthritis Outcome Score (KOOS); reliability and validity in competitive athletes after anterior cruciate ligament reconstruction

M. Safari, B. Alibhai, T. Mohammed, M. Masharani, M. Kharravi


Reliability, Validity, and Responsiveness of the IKDC Score for Menisic Injuries of the Knee


Reliability: Assignment

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reliability (internal consistency, test-retest)</th>
<th>Low-Acceptable</th>
<th>Acceptable</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOOS</td>
<td>α=0.25-0.75 r=0.61-0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IKDC</td>
<td>α=0.77-0.97 r=0.87-0.99</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Saith, 2012; Collins, 2001; Wright, 2010)
Validity: Relevance to PROs

Can we draw valid inferences from an instrument?

Gather evidence that instrument evaluates intended constructs

Related to specific purpose and setting

Content Criterion Construct

Validity: Assignment

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reliability (internal consistency, test-retest)</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOOS</td>
<td>α=0.25-0.75, r=0.61-0.95</td>
<td>Adequate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IKDC</td>
<td>α=0.77-0.97, r=0.87-0.99</td>
<td>Adequate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*A Note that sometimes information is unknown*

(Smith, 2012; Collins, 2011; Wright, 2010)

Validation

**Content**
- Extent that domain of interest is comprehensibly sampled
- Expert and patient opinion

**Criterion**
- Compare instrument against “gold standard”
- Short compared to longer instruments

**Construct**
- Evaluation of mini-theory used to explain relationship among behaviors and attitudes being studied

*Will the instrument scores allow us to draw inferences about the people that we wish to make?*
**Responsiveness**

- Effect Size
- Standardized Response Mean
- Guyatt’s Responsiveness Index
- Minimally clinically important difference (MCID)
- Clinical Important Difference (CID)
- Minimal Perceptible Clinical Improvement (MPCI)

**Power to detect true change over time**

**Clinically Meaningful**

**Why Clinical Meaningful Change?**

- Ties magnitude of change to treatment decisions
- Emphasizes primacy of patient perspective
- Easily understood.

**Interpretability**

- When has patient had meaningful change?
- Differentiate between treatment effects that are and are not beneficial?
- HRQOL score low enough to indicate impact on health status?
- Is patient better, same, or worse?
- Treatment benefit large enough to justify change in activity or management?
Interpretability: Key Values

<table>
<thead>
<tr>
<th></th>
<th>MCID</th>
<th>SEM</th>
<th>MDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Smallest amount of change patient perceives as beneficial</td>
<td>• Error: single days score</td>
<td>• Error: 2 days scores</td>
<td></td>
</tr>
</tbody>
</table>

Responsiveness and Interpretability: Assignment

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Reliability (internal consistency, test-retest)</th>
<th>Validity</th>
<th>Responsiveness</th>
<th>Error (MDC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOOS</td>
<td>α=0.25-0.75 r=0.61-0.95</td>
<td>Face: Y</td>
<td>8-10 points</td>
<td>Varies by subscale 6-12 points</td>
</tr>
<tr>
<td>IKDC</td>
<td>α=0.77-0.97 r=0.87-0.99</td>
<td>Face: Y</td>
<td>11.5 points MDC</td>
<td>9 points</td>
</tr>
</tbody>
</table>

*Minimal Perceptible Clinical Improvement

Patient needs to experience a 10 or 12 point change on the KOOS or IKDC, respectively, to be confident meaningful change has occurred.

(Smith, 2012; Collins, 2011; Wright, 2010; Roos and Lohmander 2003)

Precision of Response Categories

• Precision refers to the types of response options in a survey.
• Balance between simplicity and quality of response

Example: Binary

1. How would you rate your overall quality of life? Poor Healthy
2. I am happy. Yes No
Precision of Response Categories

Example: Visual Analogue Scale

How severe is your pain today? Place a vertical mark on the line below to indicate how bad you feel your pain is today.

No pain [ ] ] ] Very severe pain

Example: Adjectival or Likert

<table>
<thead>
<tr>
<th>Question</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Since your arm injury does it take you longer to warm up before throwing?</td>
<td>1-5</td>
</tr>
<tr>
<td>2. Has your throwing accuracy decreased since your arm injury?</td>
<td>1-5</td>
</tr>
<tr>
<td>3. Do you have pain in your arm during 75-90% effort throwing?</td>
<td>1-5</td>
</tr>
<tr>
<td>4. Has your arm injury limited your ability to ‘long toss’?</td>
<td>1-5</td>
</tr>
</tbody>
</table>

Example: Likert or Adjectival Rating

<table>
<thead>
<tr>
<th>Question</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example: Visual Analogue Scale</td>
<td></td>
</tr>
<tr>
<td>How severe is your pain today?</td>
<td></td>
</tr>
<tr>
<td>Place a vertical mark on the line below to indicate how bad you feel your pain is today.</td>
<td></td>
</tr>
<tr>
<td>No pain [ ] ] ] Very severe pain</td>
<td></td>
</tr>
</tbody>
</table>

Criteria for Selecting PRO measures

1. Instrument Development
2. Reliability
3. Validity
4. Responsiveness
5. Interpretability
6. Precision
7. Acceptability
8. Feasibility
9. Appropriateness

Instrument Essentials

Clinical Utility

• Both use Likert scales.
• 11 point scale may be more favorable depending on population.

(Smith, 2012; Collins, 2011; Wright, 2010)
Acceptability: Patient-friendliness

Number of items

Goals:
Maximize response rate & attain quality info

Criteria:
- Memorable response rate
- Quality info

An important consideration:
Finding the right balance

Number of items (and time)

Amount of information

An important consideration:
Finding the right balance

Single-item measure

Global Rating - Function
Please complete the following sentence:

On a scale of 0 to 10, my injured body part is ________ out of 10.
(0 = no pain, 10 = worst pain imaginable)

Multi-item measure

<table>
<thead>
<tr>
<th>Item</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>2</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>3</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>4</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>5</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>6</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>

© Valier and Lam, 2013

A.T. Still University
Arizona School of Health Sciences
ATSU
Acceptability: Patient-friendliness

- Number of items
- Potential comfort issues
- Time to complete
- Easily understood

Goals:
- Maximize response rate & attain quality info

(Fitzpatrick, 1998)

Acceptability: Assignment

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Number of items</th>
<th>Time to complete</th>
<th>Readability</th>
<th>Comfort level issues?</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOOS</td>
<td>42 items</td>
<td>10 minutes</td>
<td>Flesch reading ease: 74.5%</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Flesch-Kincaid readability: 4th grade</td>
<td></td>
</tr>
<tr>
<td>IKDC</td>
<td>18 items</td>
<td>5-7 minutes</td>
<td>Flesch reading ease: 64.7%</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Flesch-Kincaid readability: 6th grade</td>
<td></td>
</tr>
</tbody>
</table>

(Reyes, 2012; Collins, 2011)

Feasibility: Clinician-friendliness

- Ease of use
- Role of clinician

Goals:
- Preserve workflow and limit clinician burden

(Fitzpatrick, 1998)
A consideration:
Timeframe for administration

• How often should the PRO measure be administered?
  – Will vary depending on the patient case
    • eg. ACL reconstruction vs. ankle sprain
  – Keep in mind the recall period of the measure and its items
    • eg. Currently, how much pain…
    • eg. In the last week, how much pain…
    • eg. In the last 4 weeks, how much pain…

Feasibility: Clinician-friendliness

Goals:
Preserve workflow & limit clinician burden

No training required; easy to explain; no supervision required
No questions for clinician to complete
Recall: past 1 week
5 minutes using associated software*
None; Freely available

Feasibility: Assignment

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Ease of use</th>
<th>Role of clinician</th>
<th>Time to score</th>
<th>Costs?</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOOS</td>
<td>No training required; easy to explain; no supervision required</td>
<td>No questions for clinician to complete</td>
<td>5 minutes using associated software*</td>
<td>None; Freely available</td>
</tr>
<tr>
<td>IKDC</td>
<td>No training required; easy to explain; no supervision required</td>
<td>No questions for clinician to complete</td>
<td>5 minutes by hand**</td>
<td>None; Freely available</td>
</tr>
</tbody>
</table>

Appropriateness: Target patient population

- Does your patient population align with the intended patient population for the measure?
  - eg. condition/injuries, age?
  - Demonstrated use in other patient populations?

- NOTE: One cannot assume that the measurement properties established in one patient population would be transferable to another patient population
  - eg. adolescent versions (KOOS Kids, Pedi-IKDC)

Appropriateness: Assignment

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Patient populations</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOOS</td>
<td>Intended for knee OA (eg, young and middle aged patients) and knee injuries that may lead to OA (eg, ACL, meniscal, and chondral injuries). Demonstrated use for ligament injuries and reconstruction, meniscal tears and surgery, cartilage lesions, osteochondritis dissecans, total knee replacements, pharmacologic therapy, and glucosamine supplementation.</td>
</tr>
<tr>
<td>IKDC</td>
<td>Intended for variety of knee injuries (eg, ligament, meniscal, articular cartilage, patellofemoral pain). Demonstrated use for ligament injuries and reconstruction, meniscal tears and surgery, cartilage lesions, osteochondritis dissecans, traumatic knee dislocation, platelet-rich plasma injections, and lateral release.</td>
</tr>
</tbody>
</table>

(Reus 1998; Irrgang, 2001; Collins, 2011)
### Appropriateness: Capturing different types of changes

#### Levels of disablement
- Impairments
- Function
- Disability

#### Health-related quality of life
- Physical
- Psychological
- Social
- Others (e.g., school, spiritual, economical)

### Appropriateness: Assignment

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Impairment</th>
<th>Function</th>
<th>Disability</th>
<th>Physical</th>
<th>Psychological</th>
<th>Social</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOOS</td>
<td>19 items</td>
<td>22 items</td>
<td>1 item</td>
<td>37 items</td>
<td>2 items</td>
<td>3 items</td>
</tr>
<tr>
<td>IKDC</td>
<td>4 items</td>
<td>13 items</td>
<td>1 item</td>
<td>17 items</td>
<td>0 items</td>
<td>1 item</td>
</tr>
</tbody>
</table>

*May want to consider the addition of a generic PRO measure for the assessment of disability and/or HRQOL*

### Appropriateness: Global patient care goals

- Determining the quality of your patient care (e.g., effectiveness)
  - Choose a measure that can be applied to a wider range of patients
- Optimizing patient-centered care
  - Choose a measure that will capture changes in the specific deficits that are meaningful and important to that specific patient
Appropriateness: Assignment

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Patient populations</th>
</tr>
</thead>
</table>
| KOOS       | *Intended for knee OA (e.g., young and middle aged patients) and knee injuries that may lead to OA (e.g., ACL, meniscal, and chondral injuries).*  
Demonstrated use for ligament injuries and reconstruction, meniscal tears and surgery, chondromalacia, osteochondritis dissecans, total knee replacements, pharmacologic therapy, and glucosamine supplementation. |
| IKDC       | *Intended for variety of knee injuries (e.g., ligament, meniscal, articular cartilage, patellofemoral pain).*  
Demonstrated use for ligament injuries and reconstruction, meniscal tears and surgery, cartilage lesions, osteochondritis dissecans, traumatic knee dislocation, platelet-rich plasma injections, and lateral release. |

(Reyes 1998; Irrgang, 2001; Collins, 2011)

Decision and justification

- Summarize your decision-making process: should be based upon case scenario

<table>
<thead>
<tr>
<th>KOOS</th>
<th>IKDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument development</td>
<td>X</td>
</tr>
<tr>
<td>Reliability</td>
<td>X</td>
</tr>
<tr>
<td>Validity</td>
<td>X</td>
</tr>
<tr>
<td>Responsiveness</td>
<td></td>
</tr>
<tr>
<td>Interpretability</td>
<td></td>
</tr>
<tr>
<td>Precision</td>
<td>X</td>
</tr>
<tr>
<td>Acceptability</td>
<td>X</td>
</tr>
<tr>
<td>Feasibility</td>
<td>X</td>
</tr>
<tr>
<td>Appropriateness</td>
<td></td>
</tr>
</tbody>
</table>

Final product: a class aggregate

© Valier and Lam, 2013
A.T. Still University Arizona School of Health Sciences
ATSU
Summary: Selecting PRO measures

- Determine if the PRO measure is a scientifically sound instrument
- Determine if the PRO measure is a viable option for your patients and your clinic
- Determine if the PRO measure will support and facilitate your efforts in achieving your specific and global patient care goals

Where to start? Helpful resources.

- Disablement models, HRQOL, and clinical outcomes assessment
  - Suk et al. Musculoskeletal Outcomes Measures and Instruments (1st or 2nd edition).
  - Kaplan, S. Outcome Measurement and Management.
- Selecting measures
- PRO measure reviews
Where to start? Types of measures.

• Developing a list of commonly utilized PRO measures
  – Specific measures: region/body part, condition (e.g., OA, asthma, fear avoidance)
    • Patient specific function scale (PSFS)**
  – Generic measures
  – Single-item measures
  – Adolescent-specific measures – both specific and generic
  – PROMIS measures (NIH)

Where to start? Different approaches.

General approach
• Which patient-rated outcome measure is most appropriate for sports-related knee injuries?

Condition-specific approach
• Which patient-rated outcome measure is most appropriate for patients following an ACL injury?

Patient-specific approach
• An 16 y/o female soccer player reports to the AT clinic one week post-op ACL reconstruction. Which patient-rated outcome measure is most appropriate for patient care?

Thank you

Alison Snyder Valier, PhD, ATC
arsnyder@atsu.edu
Kenny Lam, ScD, ATC
klam@atsu.edu