Lower extremity-specific measures of disability and outcomes in orthopaedic surgery

Matthew V. Smith  
*Washington University School of Medicine in St. Louis*

Sandra E. Klein  
*Washington University School of Medicine in St. Louis*

John C. Clohisy  
*Washington University School of Medicine in St. Louis*

Geneva R. Baca  
*Washington University School of Medicine in St. Louis*

Robert H. Brophy  
*Washington University School of Medicine in St. Louis*

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.wustl.edu/open_access_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

Part of the *Medicine and Health Sciences Commons*

Please let us know how this document benefits you.

**Recommended Citation**

[https://digitalcommons.wustl.edu/open_access_pubs/1148](https://digitalcommons.wustl.edu/open_access_pubs/1148)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact [vanam@wustl.edu](mailto:vanam@wustl.edu).
Authors
Matthew V. Smith, Sandra E. Klein, John C. Clohisy, Geneva R. Baca, Robert H. Brophy, and Rick W. Wright
CURRENT CONCEPTS REVIEW

Lower Extremity-Specific Measures of Disability and Outcomes in Orthopaedic Surgery

Matthew V. Smith, MD, Sandra E. Klein, MD, John C. Clohisy, MD, Geneva R. Baca, BA, Robert H. Brophy, MD, and Rick W. Wright, MD

Investigation performed at the Department of Orthopaedic Surgery, Washington University School of Medicine, St. Louis, Missouri

Outcome measures may be simple questions or complex measures that assess multiple interrelated domains affecting treatment outcomes.

Outcome measures should be relevant to patients, easy to use, reliable, valid, and responsive to clinical changes.

Joint and disease-specific outcome measures have been developed for the hip, knee, and foot and ankle. Many of these measures would benefit from further research into their validity, reliability, and optimal applicability.

General health measures and activity level scores should be included in outcome assessments after treatment for orthopaedic conditions.

When outcome measures were reported in the early orthopaedic literature, simple metrics such as return to work and patient satisfaction were used to judge the benefit of intervention for patients. Although these simple metrics continue to be frequently used and are useful, more complex scoring systems that assess multiple interrelated domains affecting outcomes, such as pain, activities of daily living, and objective physical examination measures, have been developed. These more complex measures may include physician-based assessments, patient-based assessments, or both. They can be joint or disease-specific or focus on general health.

In general, outcome measures should be relevant to patients, easy to use, reliable, valid, and responsive to clinical changes. A reliable outcome measure consistently gives the same results under the same testing conditions. Establishing validity for an outcome measure is a complex, multifactorial task that includes evaluation of several facets, including construct validity (Does it measure what it intends to?), criterion validity (Does it correlate to other valid measures?), and content validity (Does it adequately assess the critical features of the problem?). Therefore, the true validity of an outcome is based on a body of evidence evaluating these many facets of validity rather than on a single definitive test. Furthermore, validating an instrument against another evaluates only criterion validity, which does not assess whether the instrument measures what it intends to or adequately assesses the critical features of the problem.

Responsiveness of an outcome measure is its ability to detect changes in a clinical condition. The minimal clinically important difference of an outcome measure is the smallest change in an outcome score that corresponds to a change in a patient’s condition. If the minimal clinically important difference for an outcome measure is 15 points and a patient records a 10-point change in that measure after treatment, the patient may not note a true change in his or her actual clinical condition. Lastly, outcome measures may have a so-called ceiling effect (an inability to differentiate relatively good or high outcomes) or a floor effect (an inability to differentiate relatively low or poor outcomes). A floor effect occurs when a

Disclosure: None of the authors received payments or services, either directly or indirectly (i.e., via his or her institution), from a third party in support of any aspect of this work. One or more of the authors, or his or her institution, has had a financial relationship, in the thirty-six months prior to submission of this work, with an entity in the biomedical arena that could be perceived to influence or have the potential to influence what is written in this work. No author has had any other relationships, or has engaged in any other activities, that could be perceived to influence or have the potential to influence what is written in this work. The complete Disclosures of Potential Conflicts of Interest submitted by authors are always provided with the online version of the article.
subject or a group of subjects performs poorly on an outcome measure at baseline. Because of this poor performance, the outcome measure may not be able to detect further deterioration in their condition as the subjects already approach the minimum score for that measure at baseline.

Our goal in this review is to outline the common outcome measures used to report lower-extremity outcomes in the orthopaedic literature. This is not a comprehensive or detailed evaluation of each measure but rather an overview of why each measure was developed, the conditions each has been reported to assess, and the minimal clinically important difference if available. In addition, general health (Short Form-36 [SF-36] and SF-12) and general orthopaedic outcome measures (Musculoskeletal Functional Assessment [MFA] and Short Musculoskeletal Functional Assessment [SMFA]) are commonly used in conjunction with specific lower-extremity outcome measures. These are summarized in Table I but are not included in the discussion. The lower-extremity outcome measures are summarized in Table II (hip), Table III (knee), and Table IV (foot and ankle).

### Measure for Knee and Hip Osteoarthritis

**Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC LK 3.0)**

The Likert 3.0 version of the WOMAC (WOMAC LK 3.0) is widely used and accepted as a disease-specific instrument for osteoarthritis in the lower extremities and for evaluating clinical outcomes after total hip replacement. This is a self-administered instrument with three subscales: pain (0 to 20 points), stiffness (0 to 8 points), and physical function (0 to 68 points). One total score (0 to 96 points) is also reported. There are forty-one items with Likert scale responses from 0 to 4 points. The scores are interpreted on a best-to-worst scale, in which lower values indicate less pain and higher function. The scores can be normalized, with 0 indicating severe symptoms and 100 indicating no symptoms and higher function.

Importantly, this normalization reverses the interpretation of the score (a higher score indicates no symptoms and higher function). The WOMAC is sensitive to change and easy to use; therefore, it is the most commonly used disease-specific outcome measure for osteoarthritis of the hip and knee.

The WOMAC has undergone rigorous validation and has been used in more than sixty languages, and it has been validated against the SF-36. Considering its focus on older patients with osteoarthritis, it may not be appropriate for use in a young and active population as it is likely to have ceiling effects in this group. The minimal clinically important difference was reported to be 12% of the baseline score or 6% of the maximum score in a study that examined rehabilitation intervention in osteoarthritis. A minimal clinically important difference of 9 to 12 points (on a scale of 0 to 100) has also been reported.

### Hip-Specific Outcome Measures

There exists an abundance of hip-specific clinical outcome measures as recently summarized by Suk et al. Below is a summary of the commonly used outcome measures for the hip.

**Harris Hip Score**

The Harris hip score was initially introduced in 1969 as a research tool to assess the clinical results of mold arthroplasty for traumatic hip arthritis. This scoring system was formulated to measure the important outcome variables applicable to different hip disorders and treatment techniques. Pain (44 points), functional capacity (47 points), deformity correction (4 points), and hip range of motion (5 points) are incorporated into the maximum 100-point scoring system. A higher score indicates higher function. Pain and functional capacity are heavily weighted. The initial description did not specify whether the questionnaire was clinician or patient-administered. This scoring system is a common outcome tool for total hip arthroplasty and hip joint preservation studies, and has been validated for arthroplasty.

---

**TABLE I General Health and General Orthopaedic Outcome Measures**

<table>
<thead>
<tr>
<th>Scale*</th>
<th>Anatomic Region</th>
<th>Measures</th>
<th>Validated</th>
<th>Responder Burden (no. of questions)</th>
<th>Target Population</th>
<th>MCID†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Clinician Patient</td>
<td>Age (yr) Disorders</td>
<td></td>
</tr>
<tr>
<td>MFA10</td>
<td>General orthopaedic</td>
<td>General health and function</td>
<td>Yes</td>
<td>None 101</td>
<td>≥18 Musculoskeletal, fractures, soft-tissue injuries, repetitive motion injuries, and osteoarthritis</td>
<td>Unknown</td>
</tr>
<tr>
<td>SMFA12</td>
<td>General orthopaedic</td>
<td>General health and function</td>
<td>Yes</td>
<td>None 46</td>
<td>≥18 Musculoskeletal disorders</td>
<td>Unknown</td>
</tr>
<tr>
<td>SF-3613</td>
<td>General health</td>
<td>General health</td>
<td>Yes</td>
<td>None 36</td>
<td>Any General health</td>
<td></td>
</tr>
<tr>
<td>SF-1214</td>
<td>General health</td>
<td>General health</td>
<td>Yes</td>
<td>None 12</td>
<td>Any General health</td>
<td></td>
</tr>
</tbody>
</table>

*MFA = Musculoskeletal Functional Assessment, SMFA = Short Musculoskeletal Functional Assessment, and SF = Short Form. †MCID = minimal clinically important difference.
patients against the SF-36, WOMAC, Patient-Specific Index (PASI) hip rating scale, and McMaster-Toronto Arthritis questionnaire (MACTAR)\(^29,33\). The minimal clinically important difference is 4 points with an 8% change in score\(^34\). The modified Harris hip score was developed in an attempt to construct an outcome tool for more active patients\(^35\). The questions on deformity correction (4 points) and the range of motion (5 points) were removed, leaving a total of 91 potential points for pain and function. This score is normalized to 100 by multiplying the raw score by 1.1. This score is most commonly reported for hip-joint preservation surgery. The modified score has not been validated. The minimal clinically important difference is unknown.

### Hip Disability and Osteoarthritis Outcome Score (HOOS)

The HOOS was developed to construct an instrument with improved responsiveness compared with the WOMAC LK 3.0\(^36\). This is a self-administered forty-item questionnaire with a 5-point Likert scale used to assess five subscales (pain, symptoms, activities of daily living, sport and recreation function, and hip-related quality of life). All questions on the WOMAC LK 3.0 are contained within the HOOS, and a WOMAC score can be calculated. The additional HOOS questions on sport and recreation function and hip-related quality of life have been shown to be more responsive, especially in patients younger than sixty-six years. Each subscale is scored separately and normalized to 100 points, with a higher score indicating higher function. This score has been validated relative to the SF-36\(^36\). The minimal clinically important difference is unknown.

### High-Activity Hip Outcome Measures Hip Outcome Score (HOS)

The HOS was developed as an outcome instrument for patients who underwent hip arthroscopy. This is a self-administered questionnaire with two separately scored subscales: activities of daily living and sports.

---

**TABLE II Hip Outcome Measures for Patients with Osteoarthritis and Patients with a High Activity Level**

<table>
<thead>
<tr>
<th>Scale*</th>
<th>Anatomic Region</th>
<th>Measures</th>
<th>Validated</th>
<th>Responder Burden (no. of questions)</th>
<th>Target Population</th>
<th>Age (yr)</th>
<th>Disorders</th>
<th>MCID†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hip outcome measures for patients with osteoarthritis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMAC(^15)</td>
<td>Hip and knee</td>
<td>Physical function, pain, and stiffness</td>
<td>Yes</td>
<td>None</td>
<td>41</td>
<td>55–71</td>
<td>Hip and knee osteoarthritis and arthroplasty outcomes</td>
<td>12% of baseline score or 6% of max. score; also 9–12 points on 0–100 scale</td>
</tr>
<tr>
<td>HHS(^32)</td>
<td>Hip</td>
<td>Pain, function, deformity, and range of motion</td>
<td>Yes</td>
<td>5</td>
<td>8</td>
<td>62–71</td>
<td>Traumatic osteoarthritis and/or disorders</td>
<td>7–9</td>
</tr>
<tr>
<td>HOOS(^36)</td>
<td>Hip</td>
<td>Pain, symptoms, activity limitations, sports and/or recreation, and hip-related quality of life</td>
<td>Yes</td>
<td>None</td>
<td>40</td>
<td>71.5</td>
<td>Primary hip osteoarthritis for total hip arthroplasty</td>
<td>Unknown</td>
</tr>
<tr>
<td>Hip outcome measures for patients with high activity level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOS(^37,38)</td>
<td>Hip</td>
<td>Physical function</td>
<td>Yes</td>
<td>None</td>
<td>19 for ADL and 9 for sports subscale‡</td>
<td>13–66</td>
<td>Labral tears and hip arthroscopy</td>
<td>6–9</td>
</tr>
<tr>
<td>MHHS(^35)</td>
<td>Hip</td>
<td>Pain and function</td>
<td>No</td>
<td>None</td>
<td>8</td>
<td>Any</td>
<td>Hip-joint preservation surgery</td>
<td>Unknown</td>
</tr>
<tr>
<td>Nonarthritic hip score(^40)</td>
<td>Hip</td>
<td>Pain, mechanical symptoms, physical function, and level of activity</td>
<td>Yes</td>
<td>None</td>
<td>20</td>
<td>33</td>
<td>Young adults with nonarthritic hip pain</td>
<td>Unknown</td>
</tr>
<tr>
<td>UCLA activity score(^41)</td>
<td>Lower extremity</td>
<td>Activity level</td>
<td>Yes</td>
<td>10</td>
<td>10</td>
<td>58.4–65.8</td>
<td>Osteoarthritis</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

*WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index, HHS = Harris hip score, HOOS = Hip Disability and Osteoarthritis Outcome Score, HOS = Hip Outcome Score, MHHS = modified Harris hip score, and UCLA = University of California at Los Angeles. †MCID = minimal clinically important difference. ‡ADL = activities of daily living.
daily living and sports. The activities of daily living subscale contains nineteen items (seventeen scored), and the sports subscale contains nine items pertaining to higher-level activities, such as those required in sports.\(^{37,38}\) The HOS is reported on a scale from 0 to 68 points or a scale from 0% to 100%, with a higher score indicating better function.\(^{38}\) In validity testing, the HOS activities of daily living and sports subscales had a high correlation to the SF-36 physical function subscale and physical component summary score but a lower correlation to the mental health subscale and mental component summary score. The minimal clinically important difference is 6 to 9 points.\(^{39}\)

**Nonarthritic Hip Score**

This self-administered questionnaire was developed for younger patients with higher demands than older patients with arthritic hip disease.\(^{40}\) There are twenty questions (Likert scale responses from 0 to 4), covering four domains, including pain, mechanical symptoms, physical function, and activity level. The ten questions measuring pain and physical function come directly from the WOMAC LK 3.0. The remaining questions focus on mechanical symptoms and levels of activity. The maximum score of 100 indicates normal hip function. The score was initially studied in a young patient cohort with an average age of thirty-three years, and was validated against the Harris hip score and the SF-12.\(^{40}\) This outcome tool is designed for patients undergoing nonarthroplasty hip surgery. The minimal clinically important difference is unknown.

**UCLA Activity Score**

The UCLA (University of California at Los Angeles) activity score was introduced as an activity measure for patients undergoing total hip replacement and surface replacement arthroplasty.\(^{41}\) This instrument evaluates patient activity with ten

---

**TABLE III Knee Outcome Measures**

<table>
<thead>
<tr>
<th>Scale*</th>
<th>Anatomic Region</th>
<th>Measures</th>
<th>Validated</th>
<th>Clinician (no. of items)</th>
<th>Patient (no. of questions)</th>
<th>Target Population Disorders*</th>
<th>MCID†</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOMAC(^{15})</td>
<td>Hip and knee</td>
<td>Physical function, pain, and stiffness</td>
<td>Yes</td>
<td>None</td>
<td>41</td>
<td>Hip and knee osteoarthritis and arthroplasty outcomes</td>
<td>12% of baseline score or 6% max. score; also 9-12 points on 0-100 scale</td>
</tr>
<tr>
<td>KOOS(^{44,45})</td>
<td>Knee</td>
<td>Pain, symptoms, activities of daily living, sport and recreation, and knee-related quality of life</td>
<td>Yes</td>
<td>None</td>
<td>42</td>
<td>Sports-related knee injuries</td>
<td>8–10</td>
</tr>
<tr>
<td>IKDC (Subjective Knee Form)</td>
<td>Knee</td>
<td>Symptoms, function, and sports activity</td>
<td>Yes</td>
<td>None</td>
<td>18</td>
<td>All knee conditions</td>
<td>6.3 at 6 mo, and 16.7 at 12 mo</td>
</tr>
<tr>
<td>American Knee Society Score(^{46})</td>
<td>Knee</td>
<td>Pain, stability, range of motion, and function</td>
<td>Yes</td>
<td>3</td>
<td>2</td>
<td>Knee replacement</td>
<td>Unknown</td>
</tr>
<tr>
<td>Lysholm Score(^{58})</td>
<td>Knee</td>
<td>Pain, instability, locking, squatting, limp, support, swelling, and stair-climbing</td>
<td>Yes</td>
<td>None</td>
<td>8</td>
<td>Knee ligament injuries</td>
<td>Unknown</td>
</tr>
<tr>
<td>Cincinnati Knee Rating System(^{65})</td>
<td>Knee</td>
<td>Symptoms, daily and sports activities, physical examination, stability, radiographs, and functional testing</td>
<td>Yes</td>
<td>None</td>
<td>13</td>
<td>ACL injuries, ACL reconstruction, and articular cartilage injury</td>
<td>14 at 6 mo, and 26 at 12 mo</td>
</tr>
<tr>
<td>ACL quality of life(^{69})</td>
<td>Knee</td>
<td>Physical complaints, work, recreation and sports competition, lifestyle, and social and emotional function</td>
<td>Yes</td>
<td>None</td>
<td>32</td>
<td>ACL injuries</td>
<td>Unknown</td>
</tr>
<tr>
<td>Tegner activity scale(^{59})</td>
<td>Knee</td>
<td>Activity level</td>
<td>No</td>
<td>None</td>
<td>1</td>
<td>All knee conditions</td>
<td>Unknown</td>
</tr>
<tr>
<td>Marx activity scale(^{70})</td>
<td>Knee</td>
<td>Activity level</td>
<td>Yes</td>
<td>None</td>
<td>4</td>
<td>All knee conditions</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

*\(^*\)WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index, KOOS = Knee Injury and Osteoarthritis Outcome Score, IKDC = International Knee Documentation Committee, and ACL = anterior cruciate ligament. †MCID = minimal clinically important difference.\(^{47}\)
Knee-Specific Outcome Measures

Knee Injury and Osteoarthritis Outcome Score (KOOS)

On the basis of the perceived limitations of the use of the WOMAC in young and active patients, the KOOS was created as an extension of the WOMAC. The KOOS is a patient-based assessment graded on a 5-point Likert scale and is used to evaluate outcomes after sports injuries. It evaluates pain, symptoms, activities of daily living, sport and recreation function, and knee-related quality of life. Like the WOMAC, each subscale is summed to give a total score. The total score is transformed to a scale from 0 to 100, with higher scores indicating better outcome. All WOMAC questions are contained within the KOOS so a WOMAC score can be calculated. The KOOS has been used for the assessment of a variety of diagnoses. The most sensitive subscales include the pain, sport and recreation, and knee-related quality-of-life domains that have the largest effect size in younger, more active patients. Although the minimal clinically important difference has not been formally assessed, it has been estimated from the WOMAC that a change of 8 to 10 points in a KOOS subscale is considered clinically important.

International Knee Documentation Committee (IKDC)

Subjective Knee Form

The IKDC was established in 1987, and the IKDC form was first published in 1993 as a standardized method to evaluate knee injuries and treatment. In 1997, the Board of the American Orthopaedic Society for Sports Medicine revised the form to become a knee-specific assessment tool rather than a disease and/or condition-specific tool. The result was an eighteen-question patient-rated form that evaluates symptoms, function, and sports activity. The raw scores are summed and transformed to a scale from 0 to 100, with higher scores representing better outcomes. The IKDC subjective form has been validated and shown to be reliable and responsive for a wide range of knee conditions. The strength of the IKDC form is that it can be used as a single form to assess any condition involving the knee and thus allow comparison between groups with different diagnoses. In a study including a wide variety of knee conditions, a minimal clinically important difference of 11.5 points on the 100-point scale was sensitive to change in a condition. The minimal clinically important difference of the IKDC subjective form after treatment of
articular cartilage injury is 6.3 at six months and 16.7 at twelve months14.

American Knee Society Score
The American Knee Society Score was developed in 1989 to assess the results of knee replacement and consists of two parts: (1) a knee score that evaluates pain, stability, and range of motion of the knee (three items) and (2) a function score that mainly assesses walking distance and stair-climbing (two items)55. Each portion is scored from 1 to 100, with higher scores indicating better outcome. It has been criticized as a surgeon-assessed score that introduces potential bias. It has been validated and noted to be responsive55,67. Although it is used traditionally for knee arthroplasty, it has also been used to assess patients with osteoarthritis who are not undergoing knee replacement. The minimal clinically important difference is unknown.

Lysholm
The Lysholm score was published in 1982 and modified in 1985 to evaluate outcomes after knee ligament surgery with an emphasis on the assessment of instability symptoms58. The Lysholm score is a questionnaire consisting of eight items: (1) pain, (2) instability, (3) locking, (4) squatting, (5) limp, (6) support, (7) swelling, and (8) stair-climbing. Each item is assigned a maximum numerical value, with pain and instability carrying the highest weight. The score is summed to give a number between 0 and 100 points58,69. The scores are arbitrarily categorized as excellent (95 to 100), good (84 to 94), fair (65 to 83), and poor (<65)69. While the Lysholm has been widely used to measure outcomes for knee ligament surgery59, it has received criticism that it functions better for patients after an anterior cruciate ligament (ACL) reconstruction than for those with other knee conditions58,62. Its validity, sensitivity, and reliability have been called into question58,61,62. In addition, the Lysholm scoring system may have a ceiling effect since scores tend to be higher compared with other knee outcome measures. The Lysholm score continues to have value, especially for comparison with older literature, given the wide use of the score. It may be of greater value when used in conjunction with other scores and with an activity scale, which may adjust for the ceiling effect. The minimal clinically important difference is unknown.

Cincinnati Knee Rating System
In 1983, the Cincinnati Knee Rating System was described to assess subjective symptoms and activity level58,64. Since its original description, it was modified to evaluate thirteen scales and then refined to include six subscales: (1) symptoms (20 points), (2) daily and sports activities (15 points), (3) physical examination findings (25 points), (4) stability (20 points), (5) radiographic findings (10 points), and (6) functional testing (10 points)58. The measure is scored on a 100-point scale, with higher scores indicating better outcomes. It has been validated to assess outcomes of ACL injury and reconstruction68. It is comprehensive and has undergone rigorous assessment66,67. The Cincinnati Knee Rating System scores may be lower compared with other outcome measures that assess the same condition66,67. This rating system is responsive to clinical change58.

The minimal clinically important difference of the Cincinnati Knee Rating System after treatment of articular cartilage injury is 14 at six months and 26 at twelve months68.

ACL Quality of Life
In 1998, this score was established as a disease-specific measure of chronic ACL deficiency69. There are thirty-two items that evaluate, with use of a visual analog scale, physical complaints, work-related concerns, recreational activities and sport participation or competition, lifestyle, and social and emotional function. The total score is transformed to a scale from 0 to 100 points with each item weighted equally, with higher scores indicating a better outcome. It has been shown to be valid, reliable, and responsive to clinical change67. The minimal clinically important difference is unknown.

Tegner Activity Scale
First published in 1985, the Tegner activity scale was designed to give an objective numerical value to assess a patient’s activity level60. On a scale from 0 to 10, a score of 0 represents disability secondary to knee problems. Scores between 1 and 5 represent activity levels consistent with work or recreational sports, ranging from sedentary jobs through heavy manual labor. Scores of >5 represent higher-level recreational and competitive sports. A score of 10 was assigned to national or international-level soccer. This scale relates activity to specific sports rather than to the specific skills required to participate in those sports. Thus, it is difficult to apply to all patients across different sports. Formal validation has not been performed. The minimal clinically important difference is unknown.

Marx Activity Scale
The Marx activity scale was developed to create a short, patient-based activity assessment that uses questions designed to assess specific functional activities70. It evaluates the patient’s level of activity in terms of running, cutting, decelerating, and pivoting. Designed to assess the patient’s highest peak activity over the last year, it consists of four questions that are scored from 0 to 4 on the basis of how often the activity is performed. Validated during its development, the scale is easy to use with minimal responder burden. The minimal clinically important difference is unknown.

Foot and Ankle-Specific Outcome Measures
Over sixty outcome and clinical rating systems have been described for conditions of the foot and ankle71. Many instruments are disease-specific and useful only in evaluating outcomes after treatment of conditions such as rheumatoid arthritis72–76, ankle instability77–83, Achilles tendon disorders84,85, arthritis78–88, and calcaneal fractures89–93. Below are the instruments specific to the foot and ankle region that are more widely applicable and more frequently used in outcomes research75,91–93.

American Orthopaedic Foot & Ankle Society (AOFAS) Scales
The AOFAS rating system was initially developed by the AOFAS to report clinical status for any foot or ankle disorder. There are
four separate instruments specific to regions of the foot and ankle: the ankle-hindfoot, midfoot, hallux metatarsophalangeal-interphalangeal, and lesser metatarsophalangeal-interphalangeal scales. The scales contain both clinician-based (range of motion, alignment, gait, and stability) as well as patient-based items (pain, function, walking distance and surfaces, and shoe wear) in three subscales with a maximum score of 100 points, with a lower score corresponding to greater disability. The AOFAS scales have shown low levels of validity when evaluated against SF-36, QALY (quality-adjusted life-year) scores, or the Foot Function Index (FFI)14–97. Despite their limitations, they remain some of the most commonly used outcome instruments for the foot and ankle98,99. Baumhauer et al. found the hallux and lesser toe instruments reliable in a group of patients with rheumatoid arthritis but did not establish validity of the scales100. The AOFAS scales have been established as responsive to change after intervention for foot and ankle conditions and are commonly used for this purpose101. The minimal clinically important difference is unknown.

**Foot and Ankle Ability Measure (FAAM)**

The FAAM was developed to address the need for a universal, validated instrument for the foot and ankle99. It is a patient-reported assessment used to evaluate outcomes and health status in patients with foot or ankle disorders99. The two subscales, activities of daily living and sports, are included with twenty-nine items, which are transformed to a score of 0 (greatest disability) to 100 (least disability). Validation studies have been performed for a general population of patients with foot and ankle disorders, a group of patients with diabetes mellitus, and athletes with chronic ankle instability99,100,102,103. The instrument is the most extensively validated foot and ankle outcome instrument available, with a minimal clinically important difference of 8 on the activities of daily living subscale and 9 on the sports subscale99. The score is sensitive to overall health status and comorbidities. However, limitations may exist in its use with higher-functioning patients because of a potential for a ceiling effect.

**Foot Function Index (FFI)**

The FFI is a patient-reported assessment tool to measure the impact of pain, disability, and activity restriction related to foot and ankle disorders in patients with rheumatoid arthritis7. The initial validation study was performed in a group of patients with rheumatoid arthritis, and further reliability testing has been performed in a general population of patients with a foot disorder7,10,104,105. The measure contains twenty-three items in three subscales that are scored to a maximum of 100 points. The items are scored on a visual analog scale, and a higher score indicates greater disability. This measure is a useful index for outcomes in patients with rheumatoid arthritis. Efforts to expand the use of the instrument have resulted in several modified versions (the FFI-5pt, FFI Revised, and FFI short form)106. While the FFI-5pt (the FFI with verbal rating scales) correlated well with the original FFI, the modified versions have not been used extensively. The minimal clinically important difference is unknown.

**Ankle Osteoarthritis Score (AOS)**

The AOS is a patient-reported, disease-specific assessment that was developed from a modification of the FFI and is used to measure pain and disability related to osteoarthritis of the ankle99,107,108. It has also been used as an outcome measure after treatment of tibial plafond fractures99,109,110. The score is composed of two subscales (pain and disability) with eighteen items scored on a 10-cm visual analog scale, with a maximum score of 100 points. A higher score indicates greater disability. Validation studies indicate high levels of test-retest reliability and criterion and construct validity compared with the SF-36 and WOMAC111. While this remains the only validated disease-specific instrument for ankle osteoarthritis, the primary limitation of the score is its limited applicability to a large number of foot and ankle disorders. The minimal clinically important difference is unknown.

**Mazur Ankle Score**

The Mazur ankle score is a clinician-based outcome instrument derived from the Harris hip-scoring system that is used to evaluate outcome after ankle arthrodesis84. Pain, function, and range-of-motion subscales are rated with a twelve-item instrument and a total possible score of 100 points. Results correlate well with successful ankle fusion on radiographs, with higher scores indicating a better outcome. Scores of >60 are associated with patient satisfaction85. This instrument has been primarily used in the evaluation of outcome after ankle arthrodesis and tibial plafond fractures112,113. It is limited in use because of the disease-specific development of the tool. The minimal clinically important difference is unknown.

**Victorian Institute of Sport Assessment-Achilles Questionnaire (VISA-A)**

The VISA-A is a patient-reported instrument that measures the clinical severity of Achilles tendinopathy80,115. The tool is composed of three subscales with eight items rating pain, activity, and functional status. The maximum score is 100 points, with a lower score indicating greater disability. The tool is useful in rating the clinical severity of Achilles tendinopathy for use in medical decision-making and has been validated against the grade of severity described by Percy and Conochie116, the grade of severity described by Curwin and Stanish117, and the Achilles tendinopathy range of severity80,117. It has not been validated in the measurement of clinical outcome after surgery. This instrument is also limited in its applicability to other disorders of the Achilles tendon. The Achilles tendon total rupture score was developed on the basis of the VISA-A to assess outcome after Achilles tendon rupture85. The minimal clinically important difference is unknown.

**Discussion**

Orthopaedic surgery has lagged behind other medical and surgical specialties in developing and using outcome measures to study factors that influence the outcomes after treatment118. In general, outcome measures for the lower extremity can be useful, but they vary widely in terms of what is known about their validity, reliability, minimal clinically important difference, and applicability.
to specific patient populations. It is important to understand how and why these instruments were developed, what they are validated to assess, and how these measures respond to clinical change to avoid misinterpreting reports of patient outcomes.

Investigators performing clinical research should choose measures that have been validated for the disease and/or joint in question. Investigators should also attempt to use instruments that have a reported minimal clinically important difference for ease of interpreting important changes in scores. Furthermore, when choosing outcome measures for clinical research, there is rarely a single most appropriate rating system. Studies should include a measure of general health and an activity scale since rating of disability and outcomes is often affected by psychological and sociological factors that are not accounted for in joint-specific measures. This provides better characterization of patient populations and screens for differences that may influence outcomes. For example, outcomes after an operative intervention in a sedentary patient with multiple medical comorbidities may be affected by health status compared with the same surgical procedure in an active and healthy person. Rigorous outcomes assessment requires a combination of general health measures, activity scales, and condition-specific measures. Lastly, investigators should strive to balance the utility of collecting and analyzing data from multiple outcome instruments with the burden to respondents when completing the forms.

In summary, there are a number of clinical outcome measures available for the lower extremity, many of which could benefit from further research into their validity, reliability, and optimal applicability. Investigators should use outcome measures that are valid, reliable, and responsive for assessing the condition being studied. In addition, reviewers and readers should critically evaluate the measures that are used in clinical studies. By increasing their familiarity with these instruments, orthopaedic surgeons are better equipped to design studies of lower-extremity disorders, interpret the data appropriately, and implement the findings into their practices on the basis of sound and informed judgment.

References


