Evidence of Validity for the Hip Outcome Score in Hip Arthroscopy

RobRoy L. Martin, Ph.D., P.T., C.S.C.S., and Marc J. Philippon, M.D.

Purpose: The purpose of this study was to provide evidence of validity for the Hip Outcome Score (HOS) as an outcome instrument in hip arthroscopy. Methods: We mailed the short form 36 (SF-36) and HOS questionnaires, as well as questionnaires regarding self-reported current activity level (normal, nearly normal, abnormal, or severely abnormal) and self-reported surgical outcome (excellent/good or fair/poor), to 337 subjects. Medical records were used to collect surgical and demographic information. Two groups were formed for those above and below the median age. Pearson correlation coefficients were used to assess the relation between the SF-36 scores and HOS scores. One-way analysis of variance was used to determine whether HOS scores differed according to current level of function, surgical outcome, and age. Results: Of the subjects, 116 (34%) returned the questionnaire material. Nine of these subjects reported having surgery after August 2003 and were excluded. The data analysis was therefore performed on 107 subjects (32%). Within this group, there were 56 female patients (52%) and 51 male patients (48%), with a mean age of 42 years (median, 44.2 years; range, 14 to 79 years; SD, 0.49). The HOS activities of daily living (ADL) and sports subscales had a high correlation to the SF-36 physical function subscale ($r = 0.86$ and $r = 0.84$, respectively) and physical component summary score ($r = 0.80$ and $r = 0.81$, respectively) and a significantly ($P < .005$) lower correlation to the mental health subscale ($r = 0.41$ and $r = 0.43$, respectively) and mental component summary score ($r = 0.17$ and $r = 0.18$, respectively). HOS ADL and sports subscale scores were significantly different based on current activity level, surgical outcome, and age ($P < .002$). Conclusions: This study provides evidence of validity for the HOS in a sample of subjects at a mean of 3 years after hip arthroscopy. As hypothesized, the HOS scores had a high correlation to measures of physical function and a lower correlation to measures of mental health. The HOS scores were different based on subjects’ reported current activity level, reported surgical outcome, and age. The results of this study support the use of the HOS ADL and sports subscales as a self-report outcome instrument for hip arthroscopy. Level of Evidence: Level III, development of diagnostic criteria in a study of nonconsecutive patients. Key Words: Hip Outcome Score—Hip arthroscopy—Validity—Outcome.

Although hip arthroscopy is well recognized for the treatment of a variety of musculoskeletal conditions, it continues to evolve as indications for surgery, surgical techniques, postoperative rehabilitation, and patient outcomes are better defined. Self-report instruments are commonly used and play an important role in outcome-related research in all areas of medicine. However, for the information obtained from these instruments to be useful, there must be evidence of validity to support their use.

A number of self-report instruments are available to assess individuals with musculoskeletal hip pathologies. Although the Harris Hip Score (HHS) has been modified and used as outcome score for hip arthroscopy, it is not truly a self-report instrument.
VALIDITY OF HIP OUTCOME SCORE IN ARTHROSCOPY

The HHS as originally described was designed to be administered by qualified health care professionals. All of the instruments currently available may be limited by the number items that assess activities that require a high level of ability. The Hip Outcome Score (HOS) is a promising new outcome instrument for hip arthroscopy. Evidence of content and construct validity for the HOS was investigated in 507 subjects with the diagnosis of a labral tear. Of these subjects, 263 (52%) underwent hip arthroscopy with a mean follow-up period of 6.7 months after surgery. The HOS was found to have good psychometric properties, including a strong relation with concurrent measures of physical function.

Although the HOS shows potential as an outcome instrument for hip arthroscopy, further validation is needed. Outcome studies at less than 2 years after surgery are rarely accepted. Therefore the HOS needs to have evidence of validity in a sample in which hip arthroscopy was performed more than 2 years previously on average. This evidence should include having a higher correlation to measures of a similar construct (convergent evidence) and a lower correlation to measures of a distinctly different construct (divergent evidence). Evidence of validity would also include the ability to distinguish between individuals who have varying levels of success related to their surgical outcome.

The purpose of this study was to provide evidence of validity for the HOS in a sample of subjects in which hip arthroscopy was performed at least 2 years previously. It was hypothesized that HOS scores would relate to measures of physical function while not unduly relating to measures of mental health. It was also hypothesized that HOS scores would be different based on reported current activity level, reported surgical outcome, and age.

METHODS

Procedure for Data Collection

Subjects who underwent hip arthroscopy by the senior author between August 2001 and August 2003 were identified. These 337 subjects were mailed a packet of questionnaire material that included the short form 36 (SF-36) and HOS questionnaires. The HOS contains 2 separately scored subscales: activities of daily living (ADL) and sports. The ADL subscale contains 19 items, 17 of which are scored, pertaining to basic daily activities. The sports subscale contains 9 items pertaining to higher-level activities, such as those required in athletics. Detailed information regarding instrument construction, item content, and scoring instructions, as well as a copy of the instrument itself, are available.

This packet of questionnaire material also included questions that asked subjects to rate their current level of function as normal, nearly normal, abnormal, or severely abnormal. In addition, they were asked to rate their surgical outcome as excellent/good or fair/poor. Medical and surgical records were reviewed to collect information on the surgical procedures performed, age, and gender. This study was approved by the institutional review board, and all subjects gave consent for participation in this study.

Data Analysis

Convergent validity evidence was examined by assessing the associations between the HOS and SF-36 physical function subscale and physical component summary score. Divergent evidence was examined by assessing the associations between the HOS and SF-36 mental health subscale and the mental health component summary score. Pearson correlation coefficients were used to assess these associations. Testing for differences in the correlation coefficients between the HOS and concurrent measures of physical function and mental health was done based on the equation of Meng et al. The value was adjusted to .005 to account for the multiple comparisons.

Testing the ability of the HOS to distinguish between individuals with different outcomes was done with 1-way analysis of variance (ANOVA) and post hoc analysis as appropriate. Analysis was done to assess whether HOS scores differed according to reported level of function (normal, nearly normal, abnormal, or severely abnormal) and reported surgical outcome (excellent/good or fair/poor). One-way ANOVA was also used to assess for a difference in HOS scores based on age. Two groups, younger and older, were formed based on those below or above the median age. For these analyses, \( \alpha \) was set at .05.

RESULTS

Of the subjects, 116 (34%) returned the questionnaire material. Of these, 9 reported having surgery after August 2003 and were excluded. The data analysis was therefore done on 107 subjects (32%). Within this group, there were 56 female patients (52%) and 51 male patients (48%), with a mean age of 42 years (median, 44.2 years; range, 14 to 79 years; SD, 14)
and mean time to follow-up of 3.1 years (range, 2 to 4.6 years; SD, 0.49). The following procedures were performed: labral debridement (89%), psoas release (64%) capsular modification (60%), ligamentum teres debridement (55%), microfracture (28%), iliotibial band release (12%), osteoplasty (9%), and labral repair (7%).

The HOS ADL and sports subscales were related to the SF-36 physical function subscale \((r = 0.86\) and \(r = 0.84\), respectively) and physical component summary score \((r = 0.80\) and \(r = 0.81\), respectively) while being less related to the SF-36 mental health subscale \((r = 0.41\) and \(r = 0.43\), respectively) and mental component summary score \((r = 0.17\) and \(r = 0.18\), respectively). Correlations between the HOS subscales and measures of physical function were significantly different from their correlation to measures of mental functioning \((P < .005)\).

The subjects rated their current level of function as follows: normal, 23% (n = 26); nearly normal, 42% (n = 45); abnormal, 24% (n = 24); and severely abnormal, 6% (n = 7). This question was not answered by 5% (n = 5). The mean scores on the HOS ADL and sports subscales for these groups are presented in Table 1. One-way ANOVA and post hoc comparison found that the scores for these four groups were significantly different from one another for both the ADL \((F_{3,95} = 100, P < .0005)\) and sports \((F_{3,83} = 118, P < .0005)\) subscales.

The analysis of outcome found that 80% of subjects \((n = 86)\) reported an excellent/good outcome and 19% \((n = 20)\) reported a fair/poor outcome. One subject had a missing response and was not included in this analysis. The mean scores on the HOS ADL and sports subscales for these two groups are presented in Table 1. Individuals who reported an excellent/good outcome scored significantly higher than individuals who reported a fair/poor outcome on both the ADL \((F_{1,100} = 19, P < .0005)\) and sports \((F_{1,87} = 14, P < .0005)\) subscales.

The HOS ADL and sports subscale scores for groups divided by those below and above the median age are presented in Table 1. The scores for the younger group were significantly higher than those for the older group on both the ADL \((F_{1,100} = 11.5, P = .001)\) and sports \((F_{1,87} = 16.1, P < .0005)\) subscales.

**DISCUSSION**

The results of this study offer evidence of usefulness for the HOS as an outcome instrument for individuals who have undergone hip arthroscopy. The four hypotheses of this study were affirmed. The HOS was found to have relatively high correlations with concurrent measures of physical function and relatively low correlations with concurrent measures of mental health. HOS scores were different between groups who reported normal, nearly normal, abnormal, and severely abnormal levels of function. HOS scores were also different between those who reported excellent/good outcomes and those who reported fair/poor outcomes. As expected, HOS scores were higher for younger individuals. This provides evidence that the HOS will perform differently depending on the individual’s reported activity level, surgical outcome, and age.

The HOS differs from many outcome instruments, and its performance can be influenced by a variety of factors. The HOS is designed to be sensitive to changes in physical function and is less responsive to changes in mental health. Its ability to discriminate between groups with different levels of function and surgical outcomes is a strength, as it can provide valuable information about the effectiveness of treatment. The HOS is a useful tool for evaluating the outcomes of hip arthroscopy and can help in making informed decisions about patient care.

---

**Table 1.** Mean HOS for ADL and Sports Subscales According to Reported Activity Level, Surgical Outcome, and Age

<table>
<thead>
<tr>
<th>Activity level*</th>
<th>ADL</th>
<th>Sports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>96 (range, 76-100; SD, 5)</td>
<td>94 (range, 78-100; SD, 7)</td>
</tr>
<tr>
<td>Nearly normal</td>
<td>89 (range, 68-100; SD, 8)</td>
<td>78 (range, 39-100; SD, 16)</td>
</tr>
<tr>
<td>Abnormal</td>
<td>64 (range, 35-90; SD, 17)</td>
<td>40 (range, 6-56; SD, 16)</td>
</tr>
<tr>
<td>Severely abnormal</td>
<td>31 (range, 19-47; SD, 10)</td>
<td>6 (range, 0-11; SD, 4)</td>
</tr>
<tr>
<td>Surgical outcome*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excellent/good</td>
<td>85 (range, 22-100; SD, 19)</td>
<td>72 (range, 0-100; SD, 29)</td>
</tr>
<tr>
<td>Fair/poor</td>
<td>63 (range, 19-88; SD, 21)</td>
<td>45 (range, 6-91; SD, 21)</td>
</tr>
<tr>
<td>Age*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below median age</td>
<td>87 (range, 53-100; SD, 13)</td>
<td>78 (range, 25-100; SD, 20)</td>
</tr>
<tr>
<td>Above median age</td>
<td>74 (range, 19-100; SD, 25)</td>
<td>55 (range, 0-100; SD, 33)</td>
</tr>
</tbody>
</table>

*P values comparing mean scores between groups were less than .05 for both the ADL and sports subscales.
such as the HHS, in that the score is based entirely on a self-report of functional status and performance. This may be a disadvantage in comparison to instruments that include objective physical examination findings or an assessment of symptoms. However, improvements in symptoms and signs do not necessarily result in improved functional status. In other body regions such as the knee, lumbar spine, and ankle, a relation between (1) signs and symptoms and (2) functional activity and performance has not been found. Generally, we find that individuals are most concerned about their functional status as determined by their ability to perform their desired life activities. If individuals are concerned about symptoms, it is usually because the symptoms limit their functional status. That is not to say that a patient’s symptoms and objective measures of signs are not important; however, it may not be appropriate to combine scores from items that assess signs and symptoms with those that assess function. This is an area that needs further study. We believe that the HOS can be used as a tool to complement additional information, such as perceived pain levels or radiographic findings. In addition, because it is an entirely self-reported instrument, the HOS may be well suited for follow-up assessment when objective information cannot be obtained.

A number of self-report evaluative instruments have been developed for individuals with hip pathology. These include the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Hip Disability and Osteoarthritis Outcome Score (HOOS), McMaster-Toronto Arthritis Patient Preference Disability Questionnaire (MACTAR), and Patient-Specific Index. Instruments developed for individuals with musculoskeletal pathologies of the hip, knee, ankle, or foot include the Lower Extremity Function Scale (LEFS) and American Academy of Orthopaedic Surgeons Lower Limb Questionnaire. None of these instruments has evidence to support their use in individuals who have undergone hip arthroscopy.

Two instruments that seem appropriate to discuss in more detail when considering the usefulness of the HOS compared with other instruments are the Nonarthritic Hip Score and the Modified Harris Hip Score (MHHS). The Nonarthritic Hip Score is a 20-item region-specific instrument developed for the young athletic population. These items are divided into 4 domains: pain (5 items), mechanical symptoms (4 items), physical function (5 items), and level of activity (6 items). Three items on the instrument assess sports-related activities. The responses to these 20 multiple-choice items are summed, and a total composite score is determined. Evidence of validity for the Nonarthritic Hip Score has been provided based on its correlation to the SF-12 and HHS. This study examined subjects with a mean age of 33 years without radiographic arthritic changes. Although the Nonarthritic Hip Score may prove to be valuable, further evidence to support its use in individuals after hip arthroscopy is needed.

The original HHS has been modified and used as an outcome instrument in hip arthroscopy. The MHHS includes a clinical assessment of pain (44 points) and function (47 points). There are no items on the MHHS that assess specific sports-related activities. The objective clinical assessment included with the MHHS may have an advantage over the HOS when the accuracy of patient response is questioned. Evidence of validity to support the use of the MHHS was determined in individuals at a mean of 25.7 months after hip arthroscopy. The MHHS was found to have a high correlation to the SF-36 physical function sub-scale and physical component summary score ($r = 0.71$ and $r = 0.85$, respectively). The MHHS was also able to distinguish between those patients who were on disability from those who were not.

The Nonarthritic Hip Score, MHHS, and HOS differ in item content and their assessment of higher-level activities (i.e., those required in athletics). Because the Nonarthritic Hip Score and MHHS have a limited assessment of sports-related activities, they may not be able to detect change in those functioning at a higher range of ability. This may be important in hip arthroscopy, where in many instances individuals are only limited in their ability to perform in athletics. Previous studies confirmed that the HOS ADL and sports subscales collected information and were potentially responsive throughout the range of ability, including the higher range of ability. This previous finding may be reflected in the results of this study, given that the ADL and sports subscales were able to detect differences between those who reported normal and nearly normal functioning. These two groups are presumptively functioning at a relatively high level of activity. Studies performing a direct comparison of the HOS to the Nonarthritic Hip Score and MHHS are needed and currently under way.

Previous studies have found that older individuals have poorer outcome with hip arthroscopy than younger individuals. The scores from an appropriate outcome instrument should reflect this difference. The HOS ADL and sports subscales performed as expected, with younger individuals scoring higher than older individuals.
There are a number of weaknesses associated with this study. The first is the number of subjects lost to follow-up. The lost follow-up data would be particularly detrimental if the purpose of this study was to define the outcome of hip arthroscopy. Having adequate representation of all groups with respect to reported functional level, surgical outcome, and age was critical to this study. Sufficient follow-up data were collected to have this adequate representation. Follow-up phone contact may have improved the rate of return but was not a part of our initial study design and therefore was not included in this study. A second weakness was the use of subjects’ self-reported information to define activity level and surgical outcome. Given where these individuals live and the distance that they would have to travel for a clinical follow-up visit, obtaining objective measures could not be incorporated into this study. Although this study offers evidence that the HOS is an appropriate outcome instrument for hip arthroscopy, further evidence is needed for score interpretation. This evidence would include studies of reliability and responsiveness.

CONCLUSIONS

This study provides evidence of validity for the HOS in a sample of subjects at a mean of 3 years after hip arthroscopy. As hypothesized, the HOS scores had a high correlation to measures of physical function and a low correlation to measures of mental health. The HOS scores were different based on subjects’ reported current activity level, reported surgical outcome, and age. The results of this study support the use of the HOS ADL and sports subscales as a self-report outcome instrument for hip arthroscopy.

REFERENCES