Minimal Clinically Important Difference and Substantial Clinical Benefit Values for the 12-Item International Hip Outcome Tool


Purpose: To define minimal clinically important difference (MCID) and substantial clinical benefit (SCB) values for the 12-item International Hip Outcome Tool (iHOT-12) in patients undergoing hip arthroscopy for intra-articular pathology. Methods: This was a retrospective review of prospectively collected data on patients who underwent hip arthroscopy. On initial assessment and follow-up between 335 and 395 days after surgery, subjects completed the iHOT-12 and a categorical self-rating of function (severely abnormal, abnormal, nearly normal, or normal). One-half the standard deviation (SD) of the change in 1-year iHOT-12 scores was used to calculate the MCID. Receiver operator characteristic analysis was performed to determine SCB values. A change in SCB value was determined based on an improvement in the categorical rating of function. Absolute postoperative SCB scores were calculated to determine scores that would be associated with normal function ratings or with abnormal or severely abnormal function ratings. Results: Of 1,034 eligible patients, 733 (71%) met the inclusion criteria. The subjects consisted of 537 female patients (73%) and 196 male patients (27%), with a mean age of 35.3 years (SD, 13 years). At a mean of 352 days (SD, 21 days) after surgery, 536 patients (73%) were in the “improved” group and 197 (27%) were in the “not improved” group. The MCID was 13 points. An SCB change score of 28 points was able to identify patients who improved with high sensitivity (0.79) and specificity (0.72). Scores of 86 points or greater and 56 points or less were the cutoff values found to identify subjects who rated their function as normal and abnormal, respectively, with high sensitivity (0.74 and 0.90, respectively) and specificity (0.82 and 0.86, respectively). Conclusions: This study provides information to help interpret iHOT-12 scores for a follow-up period ranging between 335 and 395 days with MCID and SCB values of 13 and 28 points, respectively. In addition, a patient who scored 86 points or better was likely to have a normal rating of function, whereas a patient with a score of 56 points or less was likely to have an abnormal rating of function. Level of Evidence: Level III, retrospective comparative study.
Patient-reported outcome measures (PROMs) continue to be emphasized when investigating the outcomes of treatment intervention for musculoskeletal conditions.\textsuperscript{1,2} To interpret score changes on a PROM when given periodically over time, there must be appropriate psychometric evidence of responsiveness. A number of PROMs have been studied in hip arthroscopy, including the 33-item International Hip Outcome Tool (iHOT-33) and 12-item International Hip Outcome Tool (iHOT-12).\textsuperscript{3,4} The iHOT-33 was developed for research purposes and has evidence to support the interpretation of scores and changes in scores over time.\textsuperscript{4-8} The iHOT-12 was developed to have similar psychometric properties to the iHOT-33 but a lower response burden and, therefore, to be more appropriate for clinical use.\textsuperscript{3} Although there is evidence of responsiveness for the iHOT-33, information is needed to meaningfully interpret changes in scores for the iHOT-12.

The iHOT-33 and iHOT-12 were designed to measure the impact of hip disease and effect of treatment in young, active individuals.\textsuperscript{3,4} One of the advantages of the International Hip Outcome Tool instruments was an extensive item development phase, which included patient involvement in both item generation and assessment of item importance. Initial testing supported use of the iHOT-33 as a valid, reliable, and responsive instrument to assess the effect of treatment in young, active patients with nonarthritic hip problems.\textsuperscript{2} The iHOT-33 was recommended for use in a research setting to assess a broad range of symptoms and problems.\textsuperscript{5} The iHOT-12 was developed to have similar characteristics of validity, reliability, and responsiveness to the iHOT-33 but with a smaller number of items to reduce patient burden and administrative effort in a clinical setting.\textsuperscript{3} It was suggested that the iHOT-12 be used for initial and postoperative follow-up assessments in routine clinical practice. Despite the iHOT-12 being only about one-third the length, Griffin et al.\textsuperscript{3} found it to be psychometrically similar to the iHOT-33. Recent studies have continued to support the use of the iHOT-33 with evidence of responsiveness and specific information that allows for interpreting scores and changes in scores.\textsuperscript{4-8} Similar evidence is not available for the iHOT-12.

Responsiveness is defined as the ability of an instrument to detect change when a change in status has occurred; it can be defined with values for the minimal clinically important difference (MCID), patient acceptable symptom state (PASS), and substantial clinical benefit (SCB).\textsuperscript{5,8,12} The MCID is defined as the smallest change in outcome that the patient is able to appreciate, whereas the PASS is defined as a satisfactory outcome status.\textsuperscript{11} SCB is considered the improvement in outcome or absolute postoperative health state that the patient considers a substantial improvement.\textsuperscript{9,10} Nwachukwu et al.\textsuperscript{8} proposed that the MCID, PASS, and SCB should be considered the lower, intermediate, and upper thresholds, respectively, for clinically significant outcome improvement. Studies have defined MCID and SCB values for the iHOT-33 but not the iHOT-12.\textsuperscript{5-8} The purpose of this study was to define MCID and SCB values for the iHOT-12 in patients undergoing hip arthroscopy for intra-articular pathology. It was hypothesized that the iHOT-12 would be responsive to change in functional status over a 1-year follow-up period for subjects who underwent hip arthroscopy.

**Methods**

This was a retrospective review of prospectively collected data maintained in a secure electronic registry. The registry consisted of patients who consented to undergo hip arthroscopy at 1 of 7 centers between January 2014 and April 2017. The inclusion criteria specific to this study included subjects who underwent surgery for intra-articular hip pathology with preoperative and follow-up iHOT-12 scores available. The follow-up data were required to be collected between 335 and 395 days after surgery. The exclusion criteria for the registry included patients with primary lumbo-pelvic pathology, advanced hip arthrosis (Tönnis grade \( > 1 \)), or other conditions contraindicated for arthroscopic hip surgery. An inability to read or understand English was also an exclusion criterion for the registry. The pre-hoc collection and storage of agreed-on clinical data points were granted according to individual institutional requirements, and institutional review board approval was granted to review the deidentified registry of patient data.

On initial assessment, subjects were given the iHOT-12 and a categorical self-rating of function to complete. For follow-up data collection, the iHOT-12 and self-rating of current function were e-mailed to subjects to complete between 335 and 395 days after surgery. Each item on the iHOT-12 was scored using a 100-mm horizontal line (visual analog scale); thus, the mean value of all questions amounts to a score result ranging from 0 to 100, with a higher score meaning better function, fewer symptoms, and higher quality of life.\textsuperscript{3} The self-rating of function consisted of the following question: “How would you rate your current level of function?” The subject chose from the following categorical responses: severely abnormal, abnormal, nearly normal, or normal. Demographic information was recorded from the electronic registry.

**Psychometric Analysis**

The MCID and SCB were calculated using distribution-based and anchor-based methods, respectively, in a similar fashion to that previously described for the iHOT-33.\textsuperscript{6-8} One-half the standard deviation
Results

Participants

Of 1,034 eligible patients, 733 (71%) met the inclusion criteria for this study and had outcome data available for analysis. The average follow-up time was 352 days (SD, 21 days). Demographic information, including, age, sex, diagnosis, and procedure performed, is presented in Table 1. It should be noted that 577 subjects (79%) underwent multiple procedures.

Psychometric Results

The mean preoperative iHOT-12, postoperative iHOT-12, and change in iHOT-12 scores were 34 points (SD, 17 points), 70 points (SD, 25 points), and 37 points (SD, 26 points), respectively. Preoperative and postoperative ratings of function are provided in Table 2. At follow-up, 536 patients (73%) were in the “improved” group and 197 (27%) were in the “not improved” group. The MCID for the change score was 13 points. The results of the ROC analysis for the change in SCB for the iHOT-12 score and absolute scores associated with a normal rating of function and with an abnormal rating of function are presented in Table 3. A change score of 28 points was able to identify patients who had a change in function versus those who did not with high sensitivity and specificity. Values of 86 points and 56 points were the cutoff scores found to be able to identify subjects who rated their function as normal and abnormal, respectively, with high sensitivity and specificity.

Discussion

This study found the iHOT-12 MCID and SCB values to be 13 and 28 points, respectively, with a score of 86 points or better identifying a normal rating of function and a score of 56 points or less identifying an abnormal rating of function. These results offer evidence of responsiveness for the iHOT-12 for a follow-up period ranging between 335 and 395 days after surgery in patients who underwent hip arthroscopy for intra-articular hip pathology. The iHOT-12 was found to have acceptable properties of responsiveness to change because the AUC value was greater than 0.80 with a 95% CI not containing 0.5. This information can be used to assess postsurgical outcomes.

Patient-centered assessments using PROMs have been increasingly used to report outcomes after orthopaedic procedures. The information obtained from a PROM is only valuable if there is evidence of reliability, validity, and responsiveness to support its use. Responsiveness may be the most relevant psychometric property to interpret changes in scores over time on a PROM. There are a number of ways to determine responsiveness. Obtaining values for the MCID and SCB as part of evidence of responsiveness has the advantage of providing clinicians with information that can directly help to interpret patient scores obtained during follow-up assessments in a clinical setting. The MCID is commonly determined using statistical methods and represents only the minimal requirement for improvement. The SCB value is an alternative psychometric value that can define improvement based on...
Table 2. Preoperative and Postoperative Rating of Function

<table>
<thead>
<tr>
<th>Rating of Function</th>
<th>Preoperatively, n</th>
<th>Follow-Up (335-395 d), n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>8 (1%)</td>
<td>224 (30%)</td>
</tr>
<tr>
<td>Nearly normal</td>
<td>138 (19%)</td>
<td>350 (48%)</td>
</tr>
<tr>
<td>Abnormal</td>
<td>461 (63%)</td>
<td>137 (19%)</td>
</tr>
<tr>
<td>Severely abnormal</td>
<td>126 (17%)</td>
<td>22 (3%)</td>
</tr>
</tbody>
</table>

on patient-centered criteria using an anchor-based question and response.6,9 When an anchor-based method is used to determine SCB values, the assessment of change is from the patient’s perspective.12 In our study, the patient-centered criterion was patients’ self-rating of current function as severely abnormal, abnormal, nearly normal, or normal. The SCB values obtained for the iHOT-12 based on a change in function rating and reported current level of function can be used as part of patient-centered assessment and may be combined with standard clinical assessment for a comprehensive evaluation of a patient’s status.

The iHOT-33 and iHOT-12 were developed as PROMs for use in young, active patients with nonarthritic hip problems.3,4 For the iHOT-33, rigorous psychometric methodology in development and patient input were used to create a comprehensive list of relevant items that assess symptoms; quality of life (social, emotional, and lifestyle concerns); and functional limitations in activities of daily living, sports or recreational activities, and work.4 The iHOT-33 has evidence of validity, reliability, and responsiveness.4-8,20 However, its length was thought to be a potential concern for clinicians who want to use a PROM for all patients at all clinical visits.3 This concern led to the development of a shorter version of the International Hip Outcome Tool. The iHOT-12 was developed by selecting 12 items from the original iHOT-33 based on psychometric evidence. Regression analysis found these 12 items accounted for greater than 96% of the total variation in the iHOT-33 score.3 In addition, the iHOT-12 scores showed excellent agreement with the iHOT-33 and almost identical sensitivity to change after treatment.3 Therefore, despite being 60% shorter, the iHOT-12 collects information similar to the original 33-item questionnaire.

A number of PROMs have been evaluated in systematic reviews for outcome assessment of patients who undergo hip arthroscopy.20-22 Kemp et al.9 evaluated the psychometric properties of the Modified Harris Hip Score, Hip Outcome Score, Hip Disability and Osteoarthritis Outcome Score, Copenhagen Hip and Groin Outcome Score, and iHOT-33 in a general hip arthroscopy population. Their study found these 5 instruments to have appropriate test-retest reliability, validity, and responsiveness, with the Hip Disability and Osteoarthritis Outcome Score and iHOT-33 performing best.7 The iHOT-12, however, was not included in the study. At a 1-year follow-up interval, studies have determined the SCB values for the iHOT-33 using ROC analysis with the anchor-based question “How would you rate your overall physical ability?” and the response choices as follows: no change, slightly worse, worse, slightly improved, or much improved.6,8 These studies have found the change score for SCB in identifying patients who were “unchanged” from “much improved” to be 24.5 points in those who underwent hip arthroscopic surgery for femoroacetabular impingement6 and 25.2 points after revision hip arthroscopy.8 These numbers are comparable with the change score of 28 points identified in our study of the iHOT-12 in a general population of patients undergoing hip arthroscopy. An absolute SCB score of 63.5 points on the iHOT-33 was able to identify patients who were much improved versus those who were unchanged among patients who underwent hip arthroscopic surgery for femoroacetabular impingement.6 This score was similar to the absolute SCB value of 56 points or less to identify an abnormal rating of function in our study. Any difference in these SCB values could be attributed to the difference in the anchor-based question and responses as well as item differences between the iHOT-12 and iHOT-33.

The iHOT-12 has been internationally recognized and translated into Dutch, German, Portuguese, Spanish, and Swedish with evidence of reliability, validity, and responsiveness to support its use.23-26 The German version of the iHOT-12 identified an important change in score of 14 points over an approximately 1-month period.23 This value was determined by asking the patients to rate the change in their hip condition over approximately 1 month as much better, somewhat better, unchanged, somewhat worse, or much worse compared with their condition at initial evaluation.23 The change in score over approximately 1 month was

Table 3. Receiver Operator Characteristic Analysis for iHOT-12 Substantial Clinical Benefit Change Score and Absolute Scores Associated With Normal and Abnormal Ratings of Function

<table>
<thead>
<tr>
<th></th>
<th>Score</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change score to identify improvement in function</td>
<td>≥28 points</td>
<td>0.79</td>
<td>0.72</td>
<td>0.82 (0.79-0.86)</td>
</tr>
<tr>
<td>Score to identify normal rating of function</td>
<td>≥86 points</td>
<td>0.74</td>
<td>0.82</td>
<td>0.87 (0.85-0.90)</td>
</tr>
<tr>
<td>Score to identify abnormal rating of function</td>
<td>≤56 points</td>
<td>0.90</td>
<td>0.86</td>
<td>0.95 (0.93-0.96)</td>
</tr>
</tbody>
</table>

AUC, area under curve; CI, confidence interval.
predictably lower than the SCB change in score of 28 points over approximately 1 year in our study.

Limitations
Although this study provides evidence of responsiveness with MCID and SCB values for the iHOT-12, there are a number of limitations. The study did not define a PASS value for the iHOT-12. The findings of our study are only applicable to patients with a follow-up period ranging between 335 and 395 days. Further studies are needed to provide specific information to help interpret scores for other evaluation periods that clinicians may typically encounter. This study is also limited to the anchor-based question and responses used herein. One of the main limitations of using anchor-based methods when determining MCID and SCB values is that the question and responses are determined by researchers and therefore do not necessarily represent what the patient considers worthwhile. Other methods to determine SCB values may provide slightly different results. This study included subjects with intra-articular hip pathology; therefore, the results cannot be used for other hip conditions. However, in most patients who undergo hip arthroscopy, multiple procedures are performed at the same time. The patients used in this data set, consisting of over 700 patients from 7 different centers across the United States, represent a fairly typical population that surgeons will encounter when using the iHOT-12. It should be noted that there was less than 80% follow-up and most subjects were female patients, which could possibly introduce bias. In relation to sex bias, Nwachukwu et al. did not show that sex influenced outcome. In addition, the values for SCB change scores may vary depending on the baseline level of function of the subjects.

Conclusions
This study provides information to help interpret iHOT-12 scores for a follow-up period ranging between 335 and 395 days with MCID and SCB values of 13 and 28 points, respectively. In addition, a patient who scored 86 points or better was likely to have a normal rating of function, whereas a patient with a score of 56 points or less was likely to have an abnormal rating of function.

References


