A Short Version of the International Hip Outcome Tool (iHOT-12) for Use in Routine Clinical Practice

Damian R. Griffin, B.M., B.Ch., M.A., M.Phil., F.R.C.S.(Tr&Orth), Nicholas Parsons, Ph.D., Nicholas G. H. Mohtadi, M.D., M.Sc., F.R.C.S.C., and Marc R. Safran, M.D., on Behalf of the Multicenter Arthroscopy of the Hip Outcomes Research Network (MAHORN)

Purpose: The purpose of this study was to develop and validate a shorter version of the 33-item International Hip Outcome Tool (iHOT-33) that could be easily used in routine clinical practice to measure both health-related quality of life and changes after treatment in young, active patients with hip disorders. Methods: A development dataset (104 patients) was explored with forward-selection linear regression analysis to choose a reduced item set for the new scale. This was tested in a validation dataset (1,833 patients) and responsiveness subset (80 patients) to measure agreement between the shorter and longer versions and to test the sensitivity of the shorter instrument to change after treatment. Results: Twelve items were chosen for a short version of the International Hip Outcome Tool (iHOT-12). The iHOT-12 showed excellent agreement with the long version (iHOT-33). It captured 95.9% (95% confidence interval, 95.0% to 96.8%) of the variation of the iHOT-33 and showed equivalent sensitivity to change with a standardized effect size of 0.98 (95% confidence interval, 0.67 to 1.28). Conclusions: A short version of the International Hip Outcome Tool (iHOT-12) has been developed. It has very similar characteristics to the original rigorously validated 33-item questionnaire, losing very little information despite being only one-third the length. It is valid, reliable, and responsive to change. We suggest that it be used for initial assessment and postoperative follow-up in routine clinical practice.

The International Hip Outcome Tool (iHOT-33) is a 33-item patient-reported measure of health-related quality of life. It was designed to measure the impact of hip disease in young, active patients and to measure the effect of treatment of this disease. Patients were extensively involved in both item genera-
tion and assessment of item importance. This patient involvement and extensive international testing during the development process led to a valid and reliable instrument for use in this particular group of patients.

The iHOT-33 includes 33 questions or items, each answered by marking a visual analog scale between 2 anchor statements. This can be done on a paper form (with a 100-mm scale) or as part of a computer-based system. The total score is calculated as a simple mean of these responses ranging from 0 to 100, with 100 representing the best possible quality-of-life score.

The iHOT-33 is most likely to be used in the research setting, for example, in randomized controlled trials to compare treatment strategies in young, active patients with hip pathologies such as femoroacetabular impingement or articular cartilage degeneration. In these studies the wide range of symptoms and problems covered by the 33 items will provide a sensitive measure of treatment-related change, and the resources associated with such studies will facilitate the use of this relatively large instrument. In routine clinical practice, most clinicians look for an instrument with 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. For example, the universal PROMs (Patient Reported Outcome Measures) program introduced by the National Health Service in England made use of the 12-item Oxford Hip Score in patients undergoing total hip arthroplasty. A similar strategy was followed for the conversion of the Short Form 36 to the shorter Short Form 12.

The purpose of this study was to develop and validate a shorter version of the International Hip Outcome Tool (iHOT) that could be easily used in routine clinical practice to measure both health-related quality of life and changes after treatment in young, active patients with hip disorders.

**METHODS**

**Development of Short Version of iHOT**

The feasibility of a short version of the iHOT was explored during a development study. During January and February 2008, active, English-speaking adults, aged 18 to 60 years, who presented as new patients to a young adult hip clinic or who were undergoing follow-up after hip-preserving treatment of hip problems were invited to take part. One hundred and four such patients completed the iHOT-33. Characteristics of these patients are shown in Table 1. A principal component analysis was used to assess the dimensionality of the iHOT-33 as a prelude to item subset selection. Eigenvalues greater than 1 were retained on the basis that any single factor should be dropped unless it contains at least as much information as any one of the original questionnaire items. Regression analysis was then used to select a reduced number of items from the 33-item set. A forward-selection procedure was used to select items that accounted for the greatest part of the variation in the overall mean of the 33 items for each patient. The variance accounted for by each item can be interpreted as a measure of how much information is captured by that item and, in a regression model, how much more information is provided by that item over that which has already been cumulatively provided by previously included items. The variance accounted for by individual regression models was assessed by the coefficient of determination, the adjusted $R^2$ value from the regression output, and expressed as a percentage where, for instance, 50% indicated that half the variance in the iHOT-33 was accounted for by the selected subset of items. This gave a rating of the importance of the items for each administration of the questionnaire. The specific ordering of the importance of individual items was viewed with some caution: As usual in such statistical modeling, the process to reach the most parsimonious linear regression model and the selection of which terms to include in that model allowed several choices. However, the results gave a strong indication as to the composition of an optimal subset of items for inclusion in a shortened form of the iHOT.

The final selection of items for inclusion in a shortened questionnaire was based on the regression analysis ranking of the relative contribution to variance of each item, a pre hoc decision to span the 4 domains of the iHOT (symptoms and functional limitations; sport and recreational activities; job-related concerns; and social, emotional, and lifestyle concerns), the item frequency-importance product, and a pragmatic intent for there to be somewhere between 10 and 15 items on the shortened questionnaire.

**Validation of Short Version of iHOT**

The shortened iHOT was validated using a separate large dataset of completed iHOT-33 questionnaires from 1,833 patients, recorded between March 2008 and September 2010. The characteristics of these patients are shown in Table 1. The mean iHOT-33 score for these patients was 44.3 (95% confidence interval [CI], 43.1 to 45.5).
For a subsample of 80 patients, both preoperative and 3-month postoperative iHOT scores were also collected. These data were used to assess responsiveness of the shortened iHOT (i.e., the sensitivity to change after treatment). Preoperative assessment of these patients was performed on the day of surgery. The median time between the preoperative and postoperative assessments for these patients was 96 days (interquartile range, 121 to 178 days).

The shortened iHOT was validated by scatterplot against the iHOT-33 and by comparison of the adjusted $R^2$ value for the validation data with that reported for the development data. A paired $t$ test was used to compare change scores (postoperative score – preoperative score) between the shortened iHOT and the iHOT-33 in the subsample of 80 patients. Responsiveness of the shortened instrument was determined using standardized effect sizes. The 104 patients from the development dataset were asked to undertake an additional administration of the iHOT-33, a mean of 24 days after the first assessment (range, 14 to 90 days); these data were used to assess test-retest reliability for the shortened instrument using an intraclass correlation coefficient.

### RESULTS

**Development**

Principal component analysis of the iHOT-33 for the development data showed that there were at least 4 important components (with eigenvalues >1) that we can loosely associate with the 4 domains of the iHOT-33. This analysis showed that there was some scope to shorten the iHOT-33 while retaining the main properties of the instrument. This was expected, because the iHOT-33 has been deliberately developed with a degree of innate redundancy to provide a measure that is both responsive to change and stable across possibly highly heterogeneous populations. Figure 1 shows the variance accounted for in the regression analysis by the inclusion of increasing numbers of items from the development data, expressed as a percentage of the total variance captured by the iHOT-33. Approximately 100% of the variance was accounted for after 20 items were included in the regression models, so the process was curtailed at this point. The very rapid rise in proportion of variance shows that a small number of items accounted for nearly all the variability in the overall mean iHOT-33 scores.

### Table 1. Patient Demographics

<table>
<thead>
<tr>
<th>Phase (No. of Patients)</th>
<th>Mean Age (yr)</th>
<th>Gender</th>
<th>Affected Hip</th>
<th>Mean Duration of Symptoms (yr)</th>
<th>Diagnoses</th>
<th>Mean Tegner Activity Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development (n = 104)</td>
<td>36.5</td>
<td>56 Male 48 Female</td>
<td>38 Left 50 Right 16 Bilateral</td>
<td>4.2</td>
<td>Chondral defects, trauma, FAI, labral tears, early OA, instability, loose bodies, Perthes disease, AVN, SCFE, dysplasia</td>
<td>6.1</td>
</tr>
<tr>
<td>Validation (n = 1,833)</td>
<td>39.3</td>
<td>1,012 Male 821 Female</td>
<td>841 Left 903 Right 89 Bilateral</td>
<td>3.6</td>
<td>Chondral defects, trauma, FAI, labral tears, early OA, instability, loose bodies, trochanteric pain, piriformis pain, synovial chondromatosis, os acetabulae, ligamentum teres tears, inflammatory arthritis, PVNS, previous osteotomy, bursitis, deformity</td>
<td>6.9</td>
</tr>
<tr>
<td>Validation: responsiveness subset (n = 80)</td>
<td>38.1</td>
<td>38 Male 42 Female</td>
<td>36 Left 41 Right 3 Bilateral</td>
<td>3.9</td>
<td>Perthes disease, AVN, SCFE, dysplasia, FAI, labral tears, early OA, trochanteric pain, bursitis, inflammatory arthritis (treated with arthroscopy, surgical hip dislocation, osteotomy, periarticular surgery)</td>
<td>5.6</td>
</tr>
</tbody>
</table>

Abbreviations: AVN, avascular necrosis; FAI, femoroacetabular impingement; OA, osteoarthritis; PVNS, pigmented villonodular synovitis; SCFE, slipped capital femoral epiphysis.
Four items, in order of importance, accounted for 99% of the variability in the overall mean of the 33 items: (1) Overall, how much pain do you have in your hip/groin? (2) How concerned are you about picking up or carrying children because of your hip? (3) How concerned are you about cutting/changing directions during your sport or recreational activities? and (4) How much trouble do you have pushing, pulling, lifting, or carrying heavy objects at work? Reassuringly, these 4 selected items represented 1 item from each of the 4 domains identified in the full iHOT-33 questionnaire.

To decide on the final selection of items, the regression analysis results, domain memberships, and frequency-importance products were considered together (Table 2). This resulted in the selection of a further 8 items to give a final selection of 12 items, detailed in Table 3, that had good frequency-importance products, covered each of the 4 domains, and accounted for greater than 99% of the variance of the iHOT-33. We called this new, shortened instrument the iHOT-12 (Appendix).

Validation

Overall iHOT-33 and iHOT-12 scores were calculated as the mean visual analog scale score for the individual items for each questionnaire for each patient. These are shown for the full validation dataset for each patient in Fig 2. There is good agreement between the 2 sets of scores, with regression analysis showing that the iHOT-12 accounted for 95.9% (95% CI, 95.0% to 96.8%) of the variation in the iHOT-33. This is close to the result of analysis of the development dataset with a value of greater than 99%.

For the subsample of patients in the validation dataset with both preoperative and postoperative scores (n = 80), change scores (postoperative score – preoperative score) were determined for both the iHOT-33 and iHOT-12. Figure 3 shows that there was excellent agreement between the iHOT-33 and iHOT-12 for these patients, with a paired t test indicating that there was no significant difference (P = 0.241) in change scores between the original questionnaire and the shortened questionnaire. Standardized effect sizes were 1.03 (95% CI, 0.70 to 1.36) and 0.98 (95% CI, 0.67 to 1.28) for the iHOT-33 and iHOT-12, respectively, indicating almost exact equivalence in responsiveness to clinical change for the 2 questionnaires. Test-retest reliability for the iHOT-12 was good, with an intraclass correlation coefficient of 0.89 (95% bootstrapped CI, 0.83 to 0.93).

### Table 2. iHOT-33 Item Domain Membership, Cumulative Variance Accounted for During Model Development Regression Analysis, and By Frequency-Importance Products

<table>
<thead>
<tr>
<th>iHOT-33 Questions</th>
<th>iHOT-33 domain</th>
<th>iHOT-33 Questions</th>
<th>iHOT-33 domain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I: Symptoms and functional limitations</td>
<td>1-16</td>
<td>I: Symptoms and functional limitations</td>
</tr>
<tr>
<td></td>
<td>II: Sports and recreational activities</td>
<td>17-22</td>
<td>II: Sports and recreational activities</td>
</tr>
<tr>
<td></td>
<td>III: Job-related concerns</td>
<td>23-26</td>
<td>III: Job-related concerns</td>
</tr>
<tr>
<td></td>
<td>IV: Social, emotional, and lifestyle</td>
<td>27-33</td>
<td>IV: Social, emotional, and lifestyle</td>
</tr>
<tr>
<td>Variance (importance)</td>
<td>80%-99% (strong)</td>
<td>16,* 21,* 23,* and 32*</td>
<td>80%-99% (strong)</td>
</tr>
<tr>
<td></td>
<td>99%-99.9% (intermediate)</td>
<td>1, 3,* 6,* 8, 11, 14,* 15, 17,* 18,* 24, 25, 26, 28,* 29,* 30, and 33*</td>
<td>99%-99.9% (intermediate)</td>
</tr>
<tr>
<td></td>
<td>&gt;99.9% (weak)</td>
<td>2, 4, 5, 7, 9, 10, 12, 13, 19, 20, 22, 27, and 31</td>
<td>&gt;99.9% (weak)</td>
</tr>
<tr>
<td>Frequency-importance product</td>
<td>&gt;60%</td>
<td>3,* 17,* 18,* 19, 20, 21,* 27, 29,* and 33*</td>
<td>&gt;60%</td>
</tr>
<tr>
<td></td>
<td>50%-60%</td>
<td>1, 2, 5, 6,* 7, 8, 9, 16,* 22, 30, and 31</td>
<td>50%-60%</td>
</tr>
<tr>
<td></td>
<td>35%-50%</td>
<td>4, 10, 11, 12, 13, 14,* 15, 23,* 24, 25, 26, 28,* and 32*</td>
<td>35%-50%</td>
</tr>
</tbody>
</table>

*Items selected for inclusion in iHOT-12.
DISCUSSION

The iHOT was developed to provide an evaluation tool for the management of nonarthritic hip problems in young, active patients. Excellent instruments already exist for patients with hip fractures, those with hip arthritis, or those undergoing hip arthroplasty. The iHOT-33 was designed using a rigorous methodology with a large number of active, young patients being considered for, or receiving, hip-preserving surgery, to capture their different problems, goals, and expectations of treatment.

The iHOT-33 is reliable; shows face, content, and construct validity; and is highly responsive to clinical change. However, it is a lengthy instrument comprising 33 separate questions in 4 domains. This is unlikely to be a problem in the context of clinical trials but might limit the usefulness of the instrument in routine clinical practice. Concerns about the practicality of the original questionnaire during pilot studies, particularly from clinicians who wanted to use it in all of their patients at first contact and on every follow-

### Table 3. Correspondence Between Items From iHOT-12 and iHOT-33

<table>
<thead>
<tr>
<th>iHOT-12 Question</th>
<th>iHOT-33 Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1: Overall, how much pain do you have in your hip/groin?</td>
<td>Question 16: How much of a distraction is your hip problem?</td>
</tr>
<tr>
<td>Question 2: How difficult is it for you to get up and down off the floor/ground?</td>
<td>Question 6: Overall, how much pain do you have in your hip/groin?</td>
</tr>
<tr>
<td>Question 3: How difficult is it for you to walk long distances?</td>
<td>Question 3: How difficult is it for you to walk long distances?</td>
</tr>
<tr>
<td>Question 4: How much trouble do you have with grinding, catching, or clicking in your hip?</td>
<td>Question 14: How much trouble do you have with grinding, catching, or clicking in your hip?</td>
</tr>
<tr>
<td>Question 5: How much trouble do you have pushing, pulling, lifting, or carrying heavy objects at work?</td>
<td>Question 23: How much trouble do you have pushing, pulling, lifting, or carrying heavy objects at work?</td>
</tr>
<tr>
<td>Question 6: How concerned are you about cutting/changing directions during your sport or recreational activities?</td>
<td>Question 21: How concerned are you about your ability to maintain your desired fitness level?</td>
</tr>
<tr>
<td>Question 7: How much pain do you experience in your hip after activity?</td>
<td>Question 17: How concerned are you about your ability to maintain your desired fitness level?</td>
</tr>
<tr>
<td>Question 8: How concerned are you about picking up or carrying children because of your hip?</td>
<td>Question 28: How much trouble do you have with sexual activity because of your hip?</td>
</tr>
<tr>
<td>Question 9: How much trouble do you have with sexual activity because of your hip?</td>
<td>Question 33: How much of the time are you aware of the disability in your hip?</td>
</tr>
<tr>
<td>Question 10: How much of the time are you aware of the disability in your hip?</td>
<td>Question 19: How much of the time are you aware of the disability in your hip?</td>
</tr>
<tr>
<td>Question 11: How concerned are you about your ability to maintain your desired fitness level?</td>
<td>Question 17: How concerned are you about your ability to maintain your desired fitness level?</td>
</tr>
<tr>
<td>Question 12: How much of a distraction is your hip problem?</td>
<td>Question 29: How much of a distraction is your hip problem?</td>
</tr>
</tbody>
</table>
up, led to a demand for a shorter version (Multicenter Arthroscopy of the Hip Outcomes Research Network, oral communication, October 2009).

The iHOT-12 uses 12 items from the original 33. Regression analysis of a development dataset identified these 12 items that accounted for greater than 99% of the total variation in the full score. In a separate, large group of patients used for a validation study, the iHOT-12 showed excellent agreement with the iHOT-33 and captured at least 96% of the variation in the full questionnaire. The iHOT-12 extends across all 4 domains identified in the work to develop the iHOT-33 and showed almost identical sensitivity to change after treatment in a mixed group of patients with a variety of pathologies and treatment modalities. Standardized effect size for both the iHOT-12 and iHOT-33 was smaller than in our previous study (around 1.0 compared with 1.8).1 We believe that this probably reflects differences in the case mix of patients presenting for hip-preserving surgery, although both patient groups had a wide variety of diagnoses and were treated with a variety of techniques.

The decision to use 12 items is, to some extent, arbitrary. As few as 4 items could be expected to capture most of the variation available with a longer questionnaire. However, it is desirable that each attribute (or dimension) that one wishes to measure has adequate representation on the questionnaire, that is, more than 1 item asking about the attribute, for 2 main reasons10: (1) multiple items decrease the variability in the overall response by increasing the measure resolution for each item, and (2) multiple items minimize the impact of idiosyncratic responses to individual items. In other words, a short questionnaire may perform well on average to discriminate between groups of patients, but it will not always capture the subtle idiosyncrasies that allow the instrument to evaluate changes in individual patients. Guyatt et al.10 recommended that 3 or 4 items should be included for each attribute. We suggest that 12 items is a reasonable compromise between a very short, simple instrument and a longer and more evaluative instrument best suited to prospective clinical research. Thus the iHOT-12 is likely to be most useful in routine clinical practice: Subtle differences within individual patients may not be identified, but this will be outweighed by the ease and speed of administration, as well as the responsiveness of the instrument on average across a practice. The iHOT-12 has excellent psychometric properties and correlates well with the iHOT-33. We suggest that the iHOT-33 will be preferred for prospective clinical studies, unless these are very large and pragmatic, where the shorter iHOT-12 may again provide an advantage.

CONCLUSIONS

A short version of the iHOT, the iHOT-12, has been developed. It has very similar characteristics to the original 33-item questionnaire, losing very little information despite being only one-third the length. It is valid, reliable, and responsive to change. We suggest that it be used for initial assessment and postoperative follow-up in routine clinical practice.


REFERENCES

INSTRUCTIONS

• These questions ask about the problems you may be experiencing in your hip, how these problems affect your life, and the emotions you may feel because of these problems.

• Please indicate the severity by marking the line below each question with a slash.

  » If you put a mark on the far left, it means that you feel you are significantly impaired. For example:

    NOT PROBLEMS AT ALL
    SIGNIFICANTLY IMPAIRED

  » If you put a mark on the far right, it means that you do not think that you have any problems with your hip. For example:

    NO PROBLEMS AT ALL
    SIGNIFICANTLY IMPAIRED

  » If the mark is placed in the middle of the line, this indicates that you are moderately disabled, or in other words, between the extremes of 'significantly impaired' and 'no problems at all'. It is important to put your mark at either end of the line if the extreme descriptions accurately reflect your situation.

• Please let your answers describe the typical situation in the last month.

---

Q1  Overall, how much pain do you have in your hip/groin?

EXTREME PAIN
NO PAIN AT ALL

Q2  How difficult is it for you to get up and down off the floor/ground?

EXTREMELY DIFFICULT
NOT DIFFICULT AT ALL

Q3  How difficult is it for you to walk long distances?

EXTREMELY DIFFICULT
NOT DIFFICULT AT ALL
Q4  How much trouble do you have with grinding, catching or clicking in your hip?

SEVERE TROUBLE  NO TROUBLE AT ALL

Q5  How much trouble do you have pushing, pulling, lifting or carrying heavy objects?

SEVERE TROUBLE  NO TROUBLE AT ALL

Q6  How concerned are you about cutting/changing directions during your sport or recreational activities?

EXTREMELY CONCERNED  NOT CONCERNED AT ALL

Q7  How much pain do you experience in your hip after activity?

EXTREME PAIN  NO PAIN AT ALL

Q8  How concerned are you about picking up or carrying children because of your hip?

EXTREMELY CONCERNED  NOT CONCERNED AT ALL

Q9  How much trouble do you have with sexual activity because of your hip?

☐ This is not relevant to me

SEVERE TROUBLE  NO TROUBLE AT ALL

Q10 How much of the time are you aware of the disability in your hip?

CONSTANTLY AWARE  NOT AWARE AT ALL

Q11 How concerned are you about your ability to maintain your desired fitness level?

EXTREMELY CONCERNED  NOT CONCERNED AT ALL

Q12 How much of a distraction is your hip problem?

EXTREME DISTRACTION  NO DISTRACTION AT ALL