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.¹ 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-

Received July 26, 2011; accepted February 26, 2012.

From Warwick Medical School, University of Warwick (D.R.G., N.P.), Coventry, England; University of Calgary Sport Medicine Centre (N.G.H.M.), Calgary, Alberta, Canada; and Stanford University (M.R.S.), Redwood City, California, U.S.A.

The authors report the following potential conflict of interest or source of funding in relation to this article: Professor Griffin is a paid consultant or employee for ConMed Linvatec and receives research or institutional support from Wright Medical. Dr. Mohtadi is on the Membership and Scientific Committees of the International Society of Arthroscopy, Knee Surgery & Orthopaedic Sports Medicine and is an Editorial Board Member of the American Journal of Sports Medicine, Clinical Journal of Sport Medicine, Physician and Sportsmedicine, and Journal of Sport-Orthopädie -Sport-Traumatologie. Dr. Safran is on the Executive Committee of the American Shoulder and Elbow Surgeons; is Treasurer and a Board Member of the International Society of Arthroscopy, Knee Surgery & Orthopaedic Sports Medicine; is Chair of the Council of Delegates and a member of the Education Committee and Board Member of the American Orthopaedic Society for Sports Medicine; is an Editorial Board Member of the American Journal of Sports Medicine; and is President and a Board Member of the Society for Tennis Medicine and Science. He receives royalites from Stryker Medical (shoulder anchor), Lippincott Williams & Wilkins, and Elsevier/Saunders; is a paid consultant or employee for ArthroCare; is an unpaid consultant for Biomimedica and Cradle Medical; receives fellowship support from Smith & Nephew, ConMed Linvatec, Ossur, and Zimmer; and owns stock or stock options in Biomimedica and Cradle Medical.

The members of the Multicenter Arthroscopy of the Hip Outcomes Research Network (MAHORN) are listed in the Acknowledgments section at the end of this article.

Address correspondence to Damian R. Griffin, B.M., B.Ch., M.A., M.Phil., F.R.C.S.(Tr&Orth), Warwick Medical School, Clifford Bridge Road, Coventry CV2 2DX, England. E-mail: damian.griffin@warwick.ac.uk

^{© 2012} by the Arthroscopy Association of North America 0749-8063/11475/\$36.00

doi:10.1016/j.arthro.2012.02.027

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 impingement² 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) program³ 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 12.4,5

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-33.¹ 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,⁶ 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 procedure⁷ 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,¹ 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).

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

TABLE 1. Patie	nt Demographics
------------------------	-----------------

Abbreviations: AVN, avascular necrosis; FAI, femoroacetabular impingement; OA, osteoarthritis; PVNS, pigmented villonodular synovitis; SCFE, slipped capital femoral epiphysis.

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.⁶ 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.



FIGURE 1. Variance (information) accounted for, expressed as a percentage, by the inclusion of increasing numbers of items in the questionnaire.

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-12.8 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 = .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

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

*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.⁹ 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.¹

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

 TABLE 3.
 Correspondence Between Items From iHOT-12 and iHOT-33

iHOT-12	Question	iHOT-33
Question 1	Overall, how much pain do you have in your hip/groin?	Question 16
Question 2	How difficult is it for you to get up and down off the floor/ground?	Question 6
Question 3	How difficult is it for you to walk long distances?	Question 3
Question 4	How much trouble do you have with grinding, catching, or clicking in your hip?	Question 14
Question 5	How much trouble do you have pushing, pulling, lifting, or carrying heavy objects at work?	Question 23
Question 6	How concerned are you about cutting/changing directions during your sport or recreational activities?	Question 21
Question 7	How much pain do you experience in your hip after activity?	Question 18
Question 8	How concerned are you about picking up or carrying children because of your hip?	Question 32
Question 9	How much trouble do you have with sexual activity because of your hip?	Question 28
Question 10	How much of the time are you aware of the disability in your hip?	Question 33
Question 11	How concerned are you about your ability to maintain your desired fitness level?	Question 17
Question 12	How much of a distraction is your hip problem?	Question 29



FIGURE 2. Relation between iHOT-33 and iHOT-12 scores for validation data (n = 1,833).

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-



FIGURE 3. Change scores (postoperative score [post-op] – preoperative score [pre-op]) for iHOT-12 and iHOT-33 (n = 80).

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.

Acknowledgment: The authors have benefited from collaboration with the Multicenter Arthroscopy of the Hip Outcomes Research Network (MAHORN), which includes J. W. Thomas Byrd, M.D.; John C. Clohisy, M.D.; Damian R. Griffin, M.A., M.Phil., F.R.C.S.(Tr&Orth); Carlos A. Guanche, M.D.; Bryan T. Kelly, M.D.; Mininder S. Kocher, M.D., M.P.H.; Christopher M. Larson, M.D.; Michael Leunig, M.D.; Hal D. Martin, D.O.; RobRoy L. Martin, Ph.D., P.T., C.S.C.S.; Joseph C. McCarthy, M.D.; Nicholas G. H. Mohtadi, M.D., M.Sc., F.R.C.S.C.; Marc J. Philippon, M.D.; Marc R. Safran, M.D.; Thomas G. Sampson, M.D.; Jon K. Sekiya, M.D.; Richard N. Villar, B.Sc., M.A., M.S., F.R.C.S; and Jason R. Werle, M.D., F.R.C.S.C. The authors are grateful to Shamumagan Karthikeyan and Jane Stokes for assistance in data collection during this study.

REFERENCES

- Mohtadi NGH, Griffin DR, Pedersen ME, et al. The development and validation of a self-administered quality-of-life outcome measure for young, active patients with symptomatic hip disease: The International Hip Outcome Tool (iHOT-33). Arthroscopy 2012;28:595-610.
- UK FASHION. A feasibility study for a randomised controlled trial comparing hip arthroscopy with best conservative care for patients with femoro-acetabular impingement. Available from: http://controlledtrials.com/ISRCTN09754699. Accessed April 1, 2012.
- Patient Reported Outcome Measures (PROMs). Available from: http://www.ic.nhs.uk/proms. Accessed June 5, 2011.
- Stewart AL, Hays RD, Ware JE. The MOS Short-form General Health Survey: Reliability and validity in a patient population. *Med Care* 1988;26:724-735.
- Ware JE, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34:220-233.
- Kaiser HF. The application of electronic computers to factor analysis. *Educ Psychol Meas* 1960;20:141-151.
- 7. Sen A, Srivastava M. Regression analysis: Theory, methods and applications. New York: Springer, 1997.
- Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: A critical review and recommendations. J Clin Epidemiol 2000;53:459-468.
- Mohtadi NG, Pedersen ME, Chan D. Assessing outcomes after hip surgery. In: Sekiya J, Safran MR, Leunig M, Ranawat A, eds. *Techniques in hip arthroscopy and joint preservation* surgery. Philadelphia: Saunders; 2011:70-80.
- Guyatt GH, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. CMAJ 1986;134:889-895.

APPENDIX

		WHICH HIP IS THIS SURVEY AU If we've asked you about one hip in pr that, Otherwise, tic which causes most
IHOT	DATE OF BIRTH	
INTERNATIONA HIP OUTCOME TOO	Today's date	Ri
QUALITY OF LIFE QUEST	IONNAIRE FOR YOUNG, ACTIVE PEOPLE WITH HIP P	ROBLEMS
INSTRUCTIONS		
 These questions ask a problems affect your 	about the problems you may be experiencing in you life, and the emotions you may feel because of thes	ir hip, how these se problems.
 Please indicate the set 	verity by marking the line below each question with	n a slash.
» If you put a mark or example:	n the far left , it means that you feel you are significan	itly impaired . For
SIGNIFICANTLY		NO PROBLEMS
If you put a mark or with your hip. For e	n the far right , it means that you do not think that you xample:	u have any problems
with your hip it of a		/
significantly Impaired		NO PROBLEMS
If the mark is placed are moderately disa 'significantly impairs your mark at either reflect your situatio	I in the middle of the line, this indicates that you bled, or in other words, between the extremes of ed' and 'no problems at all'. It is important to put end of the line if the extreme descriptions accurately n.	TIP If you don't do an activity, imagine how your hip would feel if you had to try it.
Please let your answe	rs describe the typical situation in the last month .	
Q1 Overall, how mu	ch pain do you have in your hip/groin?	
EXTREME PAIN		NO PAIN _ AT ALL
Q2 How difficult is i	t for you to get up and down off the floor/ground	d?
EXTREMELY DIFFICULT		NOT DIFFICULT AT ALL
Q3 How difficult is i	t for you to walk long distances?	
EXTREMELY		NOT DIFFICULT

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 carryir objects?	ng heavy
	SEVERE TROUBLE	NO TROUBLE AT ALL
Q6	How concerned are you about cutting/changing directions durin recreational activities?	g your sport or
	EXTREMELY CONCERNED	NOT CONCERNED AT ALL
Q7	How much pain do you experience in your hip <i>after</i> activity?	
	EXTREME PAIN	NO PAIN AT ALL
Q8	How concerned are you about picking up or carrying children be hip?	cause of your
	EXTREMELY Concerned	NOT CONCERNED AT ALL
Q9	How much trouble do you have with sexual activity because of y	our hip?
	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 desir	ed fitness level?
	EXTREMELY CONCERNED	NOT CONCERNED AT ALL
Q12	How much of a distraction is your hip problem?	
	EXTREME DISTRACTION	NO DISTRACTION