EXTENDED REPORT

Comparison of the responsiveness of the Harris Hip Score with generic measures for hip function in osteoarthritis of the hip

H L Hoeksma, C H M Van den Ende, H K Ronday, A Heering, F C Breedveld, J Dekker

....

See end of article for authors' affiliations
....

Correspondence to:
Dr H L Hoeksma; Leyweg, 275 The Hague, The Netherlands 2545ch; hlhoeksma@hetnet.nl

Accepted 3 March 2003
....

Objective: To compare responsiveness of the Harris Hip Score with generic measures (that is, the Short Form-36 (SF-36), and a test of walking speed and pain during walking) in patients with osteoarthritis (OA) of the hip.
Method: The first 75 cases within the population of a randomised clinical trial on manual therapy and exercise therapy were selected for secondary analysis. Experienced (self reported) recovery by the patients after treatment (five weeks) was used as an external criterion for clinically relevant improvement. Responsiveness was evaluated by comparing responsiveness ratios and receiver operating characteristic curves.
Results: The responsiveness ratio for the Harris Hip Score was high (1.70) compared with walking speed (0.45), pain during walking (0.66), and the subscales of the SF-36—"bodily pain" (0.42) and "physical functioning" (0.36). The area under the curve also was highest for the Harris Hip Score (0.92) compared with walking speed (0.71), pain during walking (0.73), and the SF-36 subscales—bodily pain and physical functioning (both 0.66).
Conclusion: The Harris Hip Score is more responsive than the test of walking speed, pain, and subscales for function of the SF-36 in patients with OA of the hip. The Harris Hip Score seems to be a suitable instrument to evaluate change in hip function in patients with OA of the hip.

Osteoarthritis (OA) often leads to pain, loss of mobility and muscle function, restriction in activities of daily living, and decreased quality of life.[1–7] Both exercise therapy and manual therapy are used in the management of OA.[5 8–11] These treatments both aim at improving functional ability.[5 6 11] Measures to evaluate the effect of these treatments on functional ability include the Harris Hip Score,[12 13] the Short Form-36 (SF-36),[12–18] and tests of walking speed and pain during walking.[12 18] The multidimensional Harris Hip Score was developed to evaluate outcome in orthopaedic surgery of the hip joint.[12 13] Some data are available on the responsiveness of the Harris Hip Score in evaluating outcome of total hip replacement. Results show higher responsiveness ratios for the Harris Hip Score than generic scales like the SF-36.[12–18] The SF-36 showed good responsiveness in trials on the effectiveness of various medical interventions such as total hip replacement, pharmacological therapy, and exercise programmes in OA.[14 16 19 20]

The question arises as to whether the Harris Hip Score is more responsive than generic measures for hip function in patients with OA of the hip receiving exercise therapy or manual therapy. In this study we aimed at comparing responsiveness of the Harris Hip Score with two subscales of the SF-36, and a test of walking speed and pain during walking in patients with OA of the hip.

METHOD
Responsiveness
Responsiveness is defined as the ability to detect clinically relevant changes over time.[21–30] The ability to detect change can be assessed by comparing change of an external criterion with change of the method of measurement.

In this study experienced (self reported) recovery of complaints (after treatment) was used as an external criterion

for the evaluation of responsiveness.[31] This external criterion was used to compare responsiveness of the Harris Hip Score, walking speed, pain during walking, and the SF-36.

Patients
The study was conducted within the framework of a randomised controlled trial on the effects of exercise therapy and manual therapy in OA of the hip in a hospital setting. The results of this study will be reported separately. In the period from September 1999 to December 2001, 109 patients with OA of the hip were randomly assigned to receive either manual therapy or exercise therapy. In both treatment groups patients were treated twice weekly for a period of five weeks—a total of nine treatments. All patients were referred by orthopaedic surgeons or rheumatologists to the physical therapy department of the hospital because of complaints due to OA of the hip.

Patients had primary OA according to the American College of Rheumatology (ACR) criteria for OA of the hip.[32] These criteria were assessed by the referring doctor. Exclusion criteria were (*a*) symptoms in both hips; (*b*) age <60 or >85 years; (*c*) severe complaints of the lower back; (*d*) severe cardiopulmonary disease; and (*e*) insufficient knowledge of the Dutch language to complete instructions and forms. The study was approved by the medical ethics committee of the hospital. At baseline, demographic variables, duration of complaints, use of drugs, and previous treatments were recorded. Data of the first 75 patients were selected for secondary analysis of responsiveness.

....

Table 1 Patient characteristics at baseline (n=75)*

	No	%
Sex "female'"	55	73
Duration of symptoms at baseline		
1 Week to 3 months	4	5
3–6 Months	7	9
6 Months to 1 year	15	20
1–5 Years	34	45
5–10 Years	10	13
Longer than 10 years	5	7
Radiological deterioration†		
0 (no OA)	7	9
1 (mild OA)	10	13
2 (moderate OA)	30	40
3 (severe OA)	28	37

*Mean (SD) age of patients 72 (6) years; †modified Kellgren and Lawrence score.

Figure 1 Receiver operating characteristics of stable patients versus patients with clinically relevant improvement for the Harris Hip Score.

Measurements

Radiological deterioration of OA was assessed by a radiologist according to a modified Kellgren and Lawrence scale following a standardised procedure.[33] Clinical assessments were performed (at baseline and after treatment) by a physical therapist. At both the baseline and post-treatment assessments, patients completed a patient's global assessment (experienced recovery) and questionnaires on functional ability, hip function, and quality of life. Furthermore, the same observer assessed joint range of motion and walking ability. General improvement experienced by the patient was assessed using a six point Likert scale ranging from "much worse" to "complete recovery".

The Harris Hip Score is a multidimensional observational assessment which contains eight items representing pain, walking function, activities of daily living, and range of motion of the hip joint.[13] Final score ranges from 100 (no disability) to 0 (maximum disability). The index consists of questions about pain and activities of daily living, referring to the previous week, and assessments of hip function (limping) and range of motion.

The SF-36 is a self administered questionnaire containing 36 items. Subscales represent physical functioning, bodily pain, general health, vitality, social functioning, and mental health. Response options include Likert scales and yes-no questions. All items refer to the previous week.[14 19] From the SF-36, the specific subscales "physical functioning" and "bodily pain" were considered clinically relevant for patients with OA of the hip treated with exercise therapy or manual therapy.

Walking ability was assessed using an 80 m walking test with seven turning points. If necessary, patients were allowed to use a walking aid. The test was set out in a corridor of the hospital with a marked distance of 10 m between turning points. Patients were instructed to walk fast but not to run. Time to complete the test was recorded (walking speed).

Finally, participants were asked to score pain experienced in the hip area during the test on a visual analogue scale (VAS). Walking speed and pain during the test were analysed separately.

Statistical analysis

At first, collected data were screened on normal distribution using normality plots, resulting in close to normal distribution of all test data. Next, change scores were computed for each of the evaluated scores by subtracting post-treatment scores (five weeks) from baseline scores. Change scores for walking speed in nine patients who failed to complete the walking test were not included.

Two methods were used to determine responsiveness: by calculating responsiveness ratios and by plotting receiver operatin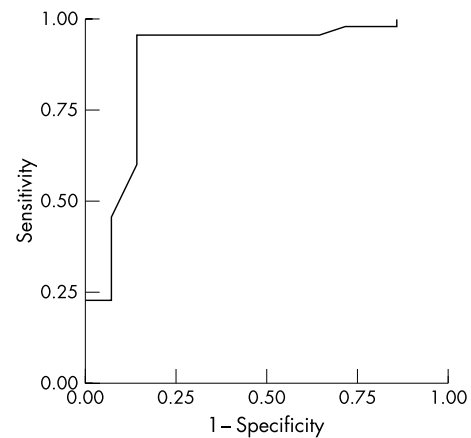g characteristic (ROC) curves. Firstly, responsiveness ratios relate a clinically relevant change to the variability of the change score in stable patients.[21–31] Secondly, ROC curves synthesise information on the sensitivity and specificity to discriminate between patients reporting clinical improvement, on the one hand, and patients reporting clinical stability on the other.[25] The point closest to the upper left corner of the curve represents the optimal trade off between sensitivity and specificity for detecting clinical improvement. The area under the curve can be interpreted as the probability of the test correctly identifying an improved patient. An area under the curve of 1.0 indicates perfect (100%) discrimination between improved and stable patients. An area under the curve of 0.5 will occur if an instrument does not discriminate between patients.[21 31 34 35]

Responsiveness ratios were calculated for the group by dividing the mean score of patients with an improved score by the standard deviation of the change in stable patients [30]. ROC curves were created by plotting the true positive proportion (sensitivity) versus the false positive proportion (1–specificity) of clinically stable patients versus clinically relevant improved patients for multiple cut off points.[21–25]

These statistics were determined for subscales of the SF-36, the Harris Hip Score, pain during the walking test, and walking speed. Data collected at baseline and after treatment (five weeks from baseline) were used. The external criterion (experienced recovery) was scored on a six point Likert scale, ranging from much worse to complete recovery. Patients scoring "much worse" and "worse" were described as having deteriorated (n=13). Patients scoring "no change" were described as clinically stable (n=14). Finally, patients scoring "improved", "much improved", and "complete recovery" were described as having clinically relevant improvement (n=48).

RESULTS

Table 1 shows the characteristics of the 75 patients included in this study. Mean age was relatively high (72 years). Most patients (58/75 (77%)) had a Kellgren and Lawrence score of 2 or 3, indicating moderate to severe OA.

Responsiveness ratios

Thirteen patients (17%) were categorised as having clinically deteriorated, 14 patients (19%) as clinically stable, and 48 (64%) as having clinically relevant improvement (table 2). The mean change scores (table 2) in clinically stable patients were close to zero, except for pain and the subscale for physical functioning of the SF-36.

The mean improvement of functional ability varied from 3.9 points on the subscale for physical functioning of the SF-36 and an improvement of 16.8 points in the Harris Hip Score,
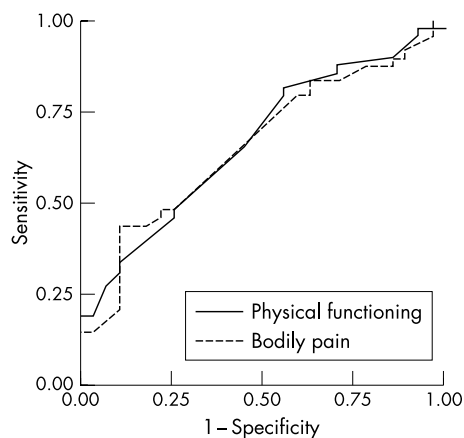
Figure 2 Receiver operating characteristics of stable patients versus patients with clinically relevant improvement for subscales of the SF-36—bodily pain and physical functioning.
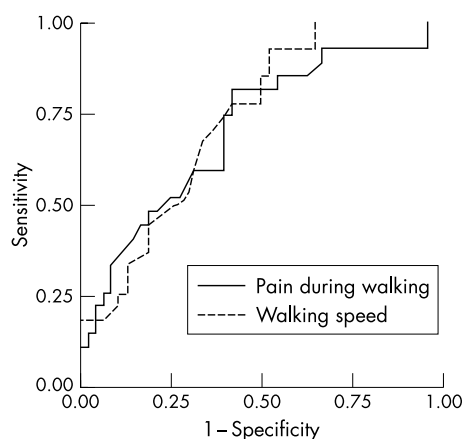

Figure 3 Receiver operating characteristics of stable patients versus patients with clinically relevant improvement for pain during walking and walking speed.

corresponding to an improvement of 10–33%. The responsiveness ratio of the Harris Hip Score was 1.70. Responsiveness ratios of walking speed (0.45), pain during walking (0.66) and the subscales of the SF-36 bodily pain (0.42) and physical functioning (0.36) were all considerably smaller.

ROC curves
Figures 1–3 present the ROC curves, showing the cut off points for the evaluated instruments.

Optimal cut off points were determined by observation of the curve coordinates and choosing the optimal trade off between sensitivity and specificity. The area under the curve of the Harris Hip Score was close to 1, implying very good sensitivity and specificity (table 3). Areas under the curve of all other measures of pain and functional ability were considerably smaller.

Table 3 Areas under the (ROC) curve and optimal cut off points

	Area under the curve	95% CI	Optimal cut off point
Harris Hip Score	0.92	0.85 to 1.00	4
SF-36 bodily pain	0.66	0.53 to 0.77	5
SF-36 physical functioning	0.66	0.53 to 0.78	3
Walking speed	0.71	0.58 to 0.83	1.5
Pain during walking	0.73	0.62 to 0.86	−3.5

DISCUSSION
We compared the responsiveness of clinical outcome measures of hip function in patients with OA of the hip. Our results suggest that the Harris Hip Score is more responsive than the physical functioning subscale of the SF-36 and a test of walking speed. Furthermore, the Harris Hip Score was more responsive than the subscale for bodily pain of the SF-36, and pain during walking. Both the ability to detect change and the ability to discriminate between improved and stable patients was highest for the Harris Hip Score, indicating that the Harris Hip score is a suitable instrument for evaluating treatment.

For the Harris Hip Score an improvement of four units, corresponding with an 8% change from baseline in our patient group, proved to be the optimal cut off point of sensitivity and specificity to detect clinical improvement. This indicates that the Harris Hip Score can detect a small improvement of 8% and higher in groups of patients.

Therefore we conclude that the Harris Hip Score is a suitable measure for evaluating the success of rehabilitation interventions, such as exercise therapy.

The multidimensional Harris Hip Score incorporates dimensions of pain, range of motion, and function. Both manual therapy and exercise therapy focus on these dimensions in patients restricted in hip function by OA. Separate analysis of the subscales of the Harris Hip Score (data not presented) indicated that functional ability as well as range of motion and pain often improved. Of the measures evaluated in this study, the Harris Hip Score combines more dimensions of functional ability than the other measures. Also, the Harris Hip Score is the only measure in this study that combines both observational and self reported items. This might be a possible explanation for the good responsiveness of the Harris Hip Score compared with other studied measures.

Angst *et al* recently compared responsiveness of the WOMAC (Western Ontario and McMaster Universities OA index) with the SF-36 in patients receiving a rehabilitation (exercise) programme.[16] The authors found that the WOMAC was more responsive than the SF-36. Both the WOMAC and the Harris Hip Scores are disease-specific measures. Like the results of Angst *et al*, our results also show that disease-specific measures are to be preferred rather than the generic SF-36 in patients with OA of the hip. A possible limitation of the study is the absence of a "gold standard" to determine clinically relevant change in patients with OA of the hip. We chose to use patient experienced recovery as an external criterion to determine clinically relevant change. Use of this method has often

Table 2 Mean change scores (with standard deviations) after five weeks (after treatment)

	No	Harris	Speed	Pain	BP	PF
Improved	48	16.8 (12.3)	−7.3 (19.9)	−13.4 (20.6)	7.6 (17.8)	3.9 (17.1)
Stable	14	0.4 (9.9)	−1.1 (16.4)	2.7 (20.4)	0.6 (18.1)	−2.1 (10.9)
Deteriorated	13	7.6 (7.9)	18.5 (30.5)	9.9 (23.2)	−5.0 (10.4)	−9.6 (12.5)

Harris, Harris Hip Score; Speed, walking speed; Pain, pain during walking; BP, SF-36 subscale for bodily pain; PF, SF-36 subscale for physical functioning. In Harris, BP, and PF, positive values indicate improvement. In Speed and Pain, negative values indicate improvement.

been reported for the responsiveness of measures.[31][35] As mentioned earlier, the use of patient experienced recovery as an external criterion to determine clinically relevant change has been criticised mainly because it may be affected by other changes in health status and subjective bias, and may not be reliable.[29][33]

However, many authors use this method in the absence of an alternative.[22][31][35] To validate our external criterion we also analysed our data with another external criterion: 20% improvement in the main complaint on a VAS. Similar results were found (data not presented).

Scoring patient's experienced recovery (a patient's global assessment) after treatment is an easy to use tool to evaluate the success of treatment, both in daily practice and in clinical research.[31][35] This might be an argument for using this measure instead of a more complex measure such as the Harris Hip Score. However, patient's experienced recovery may be influenced by subjective bias and other health concerns.[29][33] The Harris Hip Score is an observational assessment, which makes it less sensitive to a patient's subjective bias (such as socially desirable answers) than experienced recovery. Finally, functional disability (that is, decrease of range of movement and hip function) can be more specifically considered with a multidimensional measure. Therefore, in our opinion, the Harris Hip Score rather than a patient's global assessment is to be preferred.

Two features of responsiveness can be distinguished—namely, the ability of a measure to detect clinically relevant change and the ability of a measure to correctly identify a patient who has improved or not improved from a random set of data.[31][35] We used the responsiveness ratio as described by Guyatt, and ROC curves to determine these features of responsiveness.[22][24][25][31] For both responsiveness ratios and ROC curves, calculated values are highly associated with the study design.[21–24][31][34][35] Therefore, calculated values are compared with each other by most authors in order to determine the most responsive measure.

In summary, this study showed that the Harris Hip Score can detect small improvements in hip function. Therefore, we recommend that the Harris Hip Score is used in exercise therapy and other rehabilitation interventions which particularly focus on the improvement of functional ability in patients with OA of the hip.

. . . .

Authors' affiliations

H L Hoeksma, C H M Van den Ende, H K Ronday, A Heering, F C Breedveld, J Dekker, Leyenburg Hospital, The Hague, The Netherlands

REFERENCES

1 **Lequesne M**, Lamotte J, Samson M. Quality of life and functional indices in osteoarthritis. J Reum Ed Fr 1993;60:23–9.
2 **Moralestorres J**, Reginister JC, Hochberg MC. Rheumatic and musculoskeletal diseases and impaired quality of life: a challenge for rheumatologists. J Rheumatol 1996;23:1–3.
3 **Dougados M**. Clinical assessment of osteoarthritis in clinical trials. Curr Opin Rheumatol 1995;7:87–8.
4 **Steultjens MP**, Dekker J, Van Baar ME, Oostendorp RA, Bijlsma JW. Muscle strength, pain and disability in patients with osteoarthritis. Clin Rehabil 2001;3:331–41.
5 **Van Baar ME**, Assendelft WJ, Dekker J, Oostendorp RA, Bijlsma JW. Effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review of randomised clinical trials. Arthritis Rheum 1999;42:1361–9.
6 **Van Baar ME**, Assendelft WJ, Dekker J, Oostendorp RA, Bijlsma JW. The effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a randomized clinical trial. J Rheumatol 1998;25:2432–9.
7 **Steultjens MP**, Dekker J, Van Baar ME, Oostendorp RA, Bijlsma JW. Range of joint motion and disability in patients with osteoarthritis of the knee or hip. Rheumatology (Oxford) 2000;39:955–61.
8 **Hochberg MC**, Altman RD, Brandt KD, Clark MC, Dieppe PA. Griffin MR, *et al.* Guidelines for the medical management of osteoarthritis. Arthritis Rheum 1995;11:1535–40.
9 **Mazieres B**, Bannwarth B, Dougados M, Lequesne M. EULAR recommendations for the management of knee osteoarthritis. Report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials. Joint Bone Spine 2001;68:231–40.
10 **Uunk W**, Dekker D, Groenewegen P. Verwijzingen van huisartsen naar fysiotherapeuten: morbiditeitsspecifieke verwijzingspercentages: basisgegevens uit de nationale studie van ziekten en verrichtingen in de huisartspraktijk (in Dutch). Utrecht: NIVEL, 1991.
11 **Hoving JL**, Heijden Van der GJMG Fysiotherapie bij heupklachten :Systematische review van klinisch onderzoek (in Dutch). NTVF 1997;1:2–7.
12 **Shields RK**, Enloe LJ, Evans RE, Smith KB, Stechel SD. Reliability, validity and responsiveness of functional tests in patients with total joint replacement. Phys Ther 1995;75:169–76.
13 **Harris H**. Traumatic arthritis of the hip after dislocation and acetabular fracture Treatment by mold arthroplasty. J Bone Joint Surg Am 1969;4:737–55.
14 **Kosinski M**, Keller SD, Ware JE, Hatoum HT, Kong SX. The SF 36 as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: relative validity of scales in relation to clinical measures of arthritis severity. Med Care 1999;37:MS23–39.
15 **Wright JG**, Young NL. A comparison of different indices of responsiveness. J Clin Epidemiol 1997;50:239–46.
16 **Angst F**, Aeschlimann A, Steiner W, Stucki G. Responsiveness of the WOMAC index as compared with the SF 36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. Ann Rheum Dis 2001;60:834–40.
17 **Bachmeier CJ**, March LM, Cross MJ, Lapsley HM, Tribe KL, Courtenay BG, *et al.* A comparison of outcomes in osteoarthritis patients undergoing total hip and knee surgery. Osteoarthritis Cartilage 2001;9:137–46.
18 **Andersson JJ**, Firschein HE, Meenan RF. Sensitivity of a health status measure to short term clinical changes in arthritis. Arthritis Rheum 1989;32:844–50.
19 **Davies GM**, Watson DJ, Bellamy N. Comparison of the responsiveness and relative effect size of the WOMAC and the SF 36 in a randomized clinical trial of osteoarthritis patients. Arthritis Care Res 1999;12:172–9.
20 **Norman GR**, Stratford P, Regehr G. Methodological problems with retrospective computation of responsiveness to change: the lesson of cronbach. J Clin Epidemiol 1997;50:869–79.
21 **Liang MH**. Evaluating measurement of responsiveness. J Rheumatol 1995;22:1191–2.
22 **Guyatt GH**. Measuring change over time: assessing the usefulness of evaluative instruments. J Chron Dis 1987;40:171–8.
23 **Altman DG**. *Practical statistics for medical research.* London: Chapman & Hall, 1999. (CRC reprint.)
24 **Guyatt GH**, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. CMAJ 1986;134:889–95.
25 **Guyatt GH**, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. J Clin Epidemiol 1989;42:403–8.
26 **Nilsdotter AK**, Roos EM, Westerlund JP, Roos HP, Lohmander S. Comparative responsiveness of measures of pain and funcion after total hip replacement. Arthritis Care Res 2001;45:258–62.
27 **Fitzpatrick R**, Ziebland S, Jenkinson, Mowat A. Importance of sensitivity to change as a criterion for selecting health status measures. Quality in Health Care 1992;1:89–93.
28 **Hays RD**, Hadorn D. Responsiveness to change: an aspect of validity, not a seperate dimension. Quality of Life Research 1992;4:73–5.
29 **Steultjens M**, Roorda L, Dekker J, Bijlsma J. Responsiveness of observational and self-reported methods for assessing disability in mobility in patients with osteoarthritis. Arthritis Care Res 2001;45:56–61.
30 **Oberg U**, Oberg B, Oberg T. Validity and reliability of a new assessment of lower extremity dysfunction. Phys Ther 1994;74:862–71.
31 **van der Windt AWM**, van der Heiden JMG, de Winter AF, Koes BW, Deville W, Bouter LM. The responsiveness of the shoulder disablilty questionnaire. Ann Rheum Dis 1998;57:82–7.
32 **Altman J**, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991;34:505–14.
33 **Ravaud P**, Dougados M. Radiographic assessment in osteoarthritis. J Rheumatol 1997;24:786–91.
34 **Bouter LM**, van Dongen MCMJ. Epidemiologisch onderzoek (in dutch) . Houten, Bohn Stafleu Van Lochem, 1991.
35 **Streiner L**, Norman G. *Health measurement scales.* 2nd ed. Oxford:Oxford University Press, 1999.