The Development and Validation of a More Discriminating Functional Hip Score for Research

Sujith Konan, MRCS · Jenni Tahmassebi, BSc · Fares S. Haddad, MCh(Orth) FRCS(Orth)

Received: 9 December 2011/Accepted: 5 July 2012/Published online: 21 September 2012
© Hospital for Special Surgery 2012

Abstract *Background:* Total hip arthroplasty (THA) is a commonly performed procedure with increasing frequency in the young adult. While most available outcome measures can document postoperative improvement in pain and function, they do not measure the ability to perform high-demand activities. *Questions/Purposes:* We present and validate a user-friendly discriminating hip scoring system (the functional hip score) for use in younger, “high-demand” patients undergoing hip arthroplasty surgery. *Methods:* We studied 38 subjects without any hip symptoms and 72 patients undergoing THA for osteoarthritis of the hip. Preprocedure and postprocedure scores were collected in the latter cohort of patients. SF-36 and WOMAC scores were used to validate our functional scoring system. The functional hip score was tested for internal consistency, reliability, and criterion validity. *Results:* The functional hip score had high test–retest reliability, internal consistency, and criterion validity. This can be used to measure functional outcome in the younger high-demand adult patient undergoing THA. *Conclusion:* Our discriminating functional hip score can reliably measure improvement in hip function in the younger high-demand adult. Current scoring systems have ceiling effects and are unable to differentiate a high performing hip replacement from the routine hip replacement. The use of functional tasks that are measured objectively allows better documentation of improvement in hip function.

Keywords outcomes · functional scores · hip outcomes · young adult hip

Introduction

Total hip arthroplasty (THA) is one of the most commonly performed operations and is increasingly offered to younger, more active patients. Clinicians worldwide use scoring systems to document and monitor outcomes after surgical procedures. Various scoring systems for THAs are in common use [2] and range from simple measures such as the documentation of return to pre-morbid activities to formal scoring systems such as the popular Harris hip score [20]. Most scoring systems reliably measure improvement in pain or ability to perform day-to-day activities [2, 4, 10, 18, 19]. However, because we increasingly operate on young active patients who want to return to high levels of activity, any scoring system for those patients should be able to differentiate the functional ability of the younger, more active patient from that of the older arthritic patient undergoing a THA.

Although the definition of “young patient” or “high-demand activity” is arbitrary, we have identified patients who are younger than the average patients with arthritic hips but are sufficiently symptomatic to require hip replacement. They also aspire to return to more activities than simple day-to-day activities and shopping. Current scoring systems do not allow this distinction of this group of patients. For example, the Harris hip score is widely used and relatively simple. It differentiates the outcome after THA in older patients [19] and has also been used for younger patients undergoing hip arthroscopy [4]. However, we found most active patients undergoing THA could carry on functions such as use of public transport and negotiating stairs despite their pain. The Oxford hip score is another sensitive outcome

Level of Evidence: Assessment of a diagnostic evaluation. Study Level II: see Levels of Evidence for complete details.

S. Konan, MRCS (✉) · J. Tahmassebi, BSc ·
F. S. Haddad, MCh(Orth) FRCS(Orth)
Trauma and Orthopaedics, University College London Hospitals,
Euston Road,
London, WC1, UK
e-mail: docsujith@yahoo.co.uk

measure used in hip arthroplasty [10, 18]. It assesses health over the previous 4 weeks. Pain and functional ability are scored using 12 questions. It is internally consistent, valid, and reproducible [9]. However, it has no objective measure of function. Scores such as the WOMAC [6] and its modification for nonarthritic patients (the nonarthritic hip score [8]) are self-administered and easy to use but lack an objectively measurable functional hip outcome. The UCLA activity score is based on a scale of 1 to 10 with 1 being no activity and 10 being involvement in impact sports [1]. It correlates with Harris hip score and SF-12 physical component [5]. Once again, it lacks an objective measure of function. The SF-36 health survey [14, 16] and the SF-12 health survey [22] have gained considerable popularity in recent years as measures of both physical and global well-being of patients. They have been used in a wide range of medical conditions and have been validated in various parts of the world. They, however, lack the ability to specifically measure hip function and record postoperative functional improvement. None of the currently available outcome tools are able to discriminate the “high performing” hip replacement from the “routine” hip replacement. Hence, they lack the ability to test differences between specific hip replacement implants or between the young active adult and the older patients. This may be particularly necessary for research purposes.

Any new scoring system should be simple to use in a hospital setting while allowing the patient to reproduce the measures at home if necessary. It should have subjective and objective outcome measures that are easily transformed to give a total score. Therefore, the aims of this study were as follows: (1) to develop and validate a discriminating functional hip score for research in patients with osteoarthritis of the hip that could be used to demonstrate functional improvement in the younger, high-demand adult patient undergoing THA, and (2) to compare the scoring system with the WOMAC score and the SF-36 scores, two commonly used scoring systems worldwide.

Patients and Materials

The study design involved two phases: the development of the functional scoring system and then its validation.

Input from three surgeons, one physiotherapist, and patient focus groups was used to formulate the final scoring system. An active young cohort of patients who were keen to achieve better functional results was reviewed in this study. We decided to use concise reproducible functional measures that could be subjectively (performers' assessment) or objectively (observed assessment) measured so the test was not restricted to

the English-speaking population. The test was not designed to measure simple outcomes such as range of motion in the hip but instead looked at tasks requiring hip strength and proprioception. To avoid bias from the person conducting the tests (in case of objective measurements), each test was also graded subjectively by the patient on a numeric 10-point (1–10) scale for pain and difficulty separately.

The functional hip score tests five tasks. These tasks were finalized over time following observations made by physiotherapists and surgeons regarding high-end tasks that younger, more active cohorts could perform. The tasks were also discussed with patient focus groups with regards to comfort and ease of performing the tasks. Each task deals with a mechanical hip function.

Task 1 is a single leg stance (1 min). The patient is instructed to stand on the affected lower limb unaided for 1 min. The total number of times the patient drops his or her foot or uses his or her hand for balance in 1 min is measured and then allocated to categories on an ordinal scale.

Task 2 is a timed stair climb (ten stairs). The patient is instructed to climb ten stairs (height) at the quickest possible pace. The time taken for this is measured and then allocated to categories on an ordinal scale.

Task 3 is a lateral step up onto stairs. The patient is instructed to stand unaided, flex his or her affected hip and knee, and place it sideways onto a comfortable step on the stair. The support used by the patient to balance this act once is measured and then allocated to categories on an ordinal scale.

Task 4 is three forward jumps, standing up between jumps. The patient flexes his or her hip to a comfortable degree, squeezes his or her hands against his or her waist, and jumps forward three times. The ease of performing this task is measured and then allocated to categories on an ordinal scale.

Task 5 is three sideways jumps, standing up between jumps. The patient flexes his or her hip to a comfortable degree, squeezes his or her hands against his or her waist, and jumps sideways three times. The ease of performing this task is measured and then allocated to categories on an ordinal scale.

Each task is scored on a mutually exclusive scale of four choices that are ordered in the same hierarchical arrangement for all tests. For each task, the patient also grades the pain associated with performing the test and the difficulty of performing the task, respectively, on a scale of 1 to 10. A value of 10 represents inability to perform the given task. All scores from the tasks were recorded and used unweighted to avoid any preconceived bias by the person interpreting the results.

The final results of the functional hip score were calculated and interpreted as sets of three: function (F), pain (P), and difficulty (D). There are five physical

tests. Each test is scored in three ways (a higher score is worse):

1. The patient's perception of pain: score 1–10 (termed P)
2. The patient's perception of difficulty: score 1–10 (termed D)
3. The rater's measurement of function: score 1–4 (termed F)

The values of P (pain) and D (difficulty) range from 5 to 50 (5×1 to 5×10) and F (function) from 5 to 20 (5×1 to 5×4). P (pain) and D (difficulty) scores are then multiplied by 2 and F (function) score by 5 so all scores end up with a highest possible score of 100. Thus, the best score achieved on the functional hip score would be F25, P10, and D10, whereas the worst score would be F100, P100, and D100.

To validate the functional hip score, two study groups were chosen. The first group had no musculoskeletal disease. These subjects were interviewed and examined by physiotherapists and gave no history or signs suggestive of hip or any other joint disease. They were members of staff working in our hospital and people who consented to try the functional tasks when the functional scoring system was first developed. There were 38 members in this group with 17 females and 11 males, and their age ranged between 30 and 50 years (average, 46 years). The second group was a consecutive series of 72 patients who presented to the senior author (F.S.H.) with osteoarthritic hip pain and underwent THA. There were 43 females and 29 males with an average age of 49 years (range, 33–55 years). There was no difference in mean age between males and females. These patients had symptomatic arthritis affecting one hip joint only. Three patients had drug-controlled hypertension, and two patients had diet-controlled diabetes. The functional hip score, the WOMAC score, and the SF-36 scores were collected from these patients preoperatively and 1 year postoperatively. Thus, two cohorts of subjects were tested for this validation study [no musculoskeletal disease, preoperative osteoarthritis, and postoperative (12 months)]: patients who had undergone THA.

Fifteen patients randomly selected using sealed envelopes from the cohort of 38 subjects with no musculoskeletal disease or other comorbidities were initially tested using the functional hip score to assess the reproducibility of the scoring system. These subjects were tested 2 weeks apart. The intraclass correlation coefficient was calculated for the correlation between the tests' scores performed at the two time points. A good correlation assumed good reproducibility.

Cronbach's alpha coefficients were computed on each aspect of the functional score to estimate their internal consistency and reliability.

To validate the ability of the functional hip score to assess the desired improvement of hip function (criterion validity), scores from pre- and postoperative groups

were compared against two well-known validated scoring systems, the SF-36 health survey and the WOMAC. Because the functional score uses three different components, the "pain" section was compared with the pain section of the WOMAC score and bodily pain section of SF-36. The "difficulty" component was compared with the combined score of the WOMAC score. The "functional" component of the functional score was compared with the role physical (PF) section of the SF-36 score. Completed SF-36 forms and WOMAC forms were obtained from all patients. The tasks of the functional hip score were also completed by all patients. The nonparametric Kendall's tau correlation coefficient was used to establish comparisons between the various aspects of the three scoring systems. A correlation coefficient of 1 would indicate very good correlation with 0 indicating no correlation and -1 indicating negative correlation.

All three scores were obtained preoperatively and 1 year postoperatively in the 72 patients who underwent THA. Correlations between the functional hip score and the SF-36 and WOMAC scores were analyzed pre- and postoperatively to document sensitivity to clinical change. Cohen's *d* and effect size correlation were calculated to demonstrate the magnitude of change in scores after intervention. Cohen's *d* was calculated as the difference in pre- and postoperative values divided by the pooled standard deviation. Higher values suggest larger magnitude of improvement from the intervention.

Persistently worse scores and inability to demonstrate improvement in symptoms by the scoring systems were studied by looking at "floor" and "ceiling" effects. The presence of more than 15% of the scores in the best or worst possible range was considered as ceiling and floor effects, respectively.

All statistical analyses were performed using the SPSS 10.1 statistical package (SPSS Inc., Chicago, IL, USA).

Results

A high test-retest reliability [intraclass correlation coefficient (ICC), 0.97; range, 0.90–0.98] was noted for the functional hip score when the tests were repeated in a cohort of 15 patients 2 weeks apart. The ICC was also high for individual tasks of the functional scoring system (0.87 for single-leg stance, 0.84 for timed stair climb, 0.89 for lateral step up, 0.88 for forward jump, and 0.90 for sideways jump). This confirms the ability to reliably reproduce the results of the functional tasks with repeated measurements.

The Cronbach's alpha was high reflecting the internal consistency of the various subsets of the functional hip score to consistently measure a similar underlying concept (Table 1). Cronbach's alpha will generally increase as the intercorrelations among test items increase. A

Table 1 Internal consistency of the individual tasks of the functional hip score

	Cronbach's alpha
No hip pathology	
Single leg stance	0.76
Timed stair climb	0.77
Lateral step up stair	0.81
Forward jump	0.83
Sideways jump	0.83
Osteoarthritis preoperative patients	
Single leg stance	0.67
Timed stair climb	0.74
Lateral step up stair	0.71
Forward jump	0.85
Sideways jump	0.77
Post-THA patients	
Single leg stance	0.76
Timed stair climb	0.74
Lateral step up stair	0.69
Forward jump	0.71
Sideways jump	0.78

high alpha noted in our study indicates that the set of tasks can measure a single construct as intercorrelations among test items are maximized when all items measure the same construct.

There was an improvement between the preoperative and postoperative scores between all three scoring systems (Table 2). All three components of the functional hip score were individually able to detect change in the functional status of the patient after the intervention (THA). This is demonstrated by the effect size, which was higher with the functional hip score and the WOMAC scores. Higher values of effect size suggest larger magnitude of improvement from the intervention (hip replacement in this case). The phenomenon of ceiling and floor effect was noted only with WOMAC and SF-36 scores. For WOMAC scores in the postoperative cohort, 38% in the pain category, 26% of the stiffness category, and 35% of the disability category reported best possible

results. Similarly, in the WOMAC preoperative cohort, 17%, 22%, and 19% of the patients reported worse possible results. Analysis of the SF-36 scores also demonstrated ceiling effects in general health (GE), social function (SF), and role emotional (RE) subscales of the postoperative cohort and floor effects in the role physical (RP) and role emotional (RE) subscales. The “pain” component of the functional hip score in the cohort of patients with no hip disease correlated with the WOMAC pain component with a Kendall's tau correlation coefficient of 0.63 and with the “body pain” component of the SF-36 score (0.43). A similar result was noted with preoperative patients ($r=0.66$ for WOMAC and 0.51 for SF-36 scores) and postoperative patients ($r=0.78$ with WOMAC and 0.46 with SF-36). The “difficulty” component of the functional hip score correlated with WOMAC score ($r=0.88$) in the cohort with no hip disease and in the pre- and postoperative cohorts (coefficient of 0.54 and 0.49, respectively). A high correlation was also noted between the “functional” component of the hip score and the SF-36 PF section in all three cohorts ($r=0.69$ in the no hip disease group, 0.67 in the preoperative group, and 0.58 in the postoperative group). A good correlation ($r=0.53$, 0.64, and 0.57) was noted between the WOMAC scores and SF-36 physical component in all three cohorts in our study.

Discussion

There is a need for a discriminating functional hip score to measure outcome in the high-demand younger adult undergoing THA. The aim of our study was to develop and validate a functional hip score which can be used in the young high-demand adult undergoing hip replacement surgery.

The development of an objective function-based outcome score has limitations. The functional hip score was developed specifically to look at high-demand patients undergoing hip arthroplasty and their return to desired level of

Table 2 Comparison between the three scoring systems

	Preoperative scores (standard deviation)	Postoperative scores (standard deviation)	Cohen's <i>d</i> (effect size correlation)	Two-tailed <i>t</i> test (<i>p</i> value)
Functional hip score				
Function	78.53 (11.58)	62.94 (15.82)	2.51 (0.78)	<0.01
Difficulty	72.11 (8.93)	46.88 (15.53)	1.99 (0.71)	<0.01
Pain	76.53 (7.44)	21.29 (12.54)	5.36 (0.94)	<0.01
WOMAC				
Pain	30.70 (8.02)	14.76 (7.26)	2.23 (0.74)	<0.01
Disability	23.31 (5.19)	12.71 (6.37)	2.42 (0.77)	<0.01
Stiffness	75.11 (17.33)	33.94 (14.90)	1.82 (0.67)	<0.01
SF-36				
Physical component	26.18 (5.92)	35.69 (9.97)	1.16 (0.50)	0.01
Mental component	42.68 (13.47)	50.91 (13.00)	0.62 (0.30)	0.01

function. The use of our scoring system in a general population undergoing hip replacement may not always be practical. We have used two commonly used scoring systems: the WOMAC scores and the SF-36 to validate the functional hip score. These are commonly used scores in our institution. This study shows that our functional score is reproducible and valid and may be used to demonstrate differences in outcomes in the younger high-demand adult or between implants. It would be interesting to test our functional scores against other available scores such as the Harris hip score and the UCLA scale. We are currently assessing the usefulness of the functional hip score in other hip interventions such as hip resurfacing, pelvic osteotomy, and hip arthroscopy. This functional score may prove particularly useful to assess outcome in a younger cohort of patients with higher activity levels such as those undergoing hip arthroscopy for impingement syndromes.

An outcome measure should be reliable, valid, and sensitive to change [2]. Repeatability of a scoring system can be measured by correlating two separate recordings of the scoring systems at two different time points. Using Kendall's tau correlation coefficient tests, the criterion validity of a scoring system is tested, and comparisons are made with the established "gold standard" in the field. We choose two widely used scoring systems for comparison, as there is no established gold standard to measure hip function post hip replacement in the young high-demand adult. To measure the ability of a scoring system to detect sensitivity to change, scores before and after an intervention are compared for statistical significance and effect size. Our scoring system showed a high correlation to the SF-36 and the WOMAC scores. The SF-36 score is a widely used generic questionnaire that is valid and consistent [12, 21], sensitive [7], and reproducible [17]. Although it is not specific to hip disease, it is valid [11] and reliable [15] in patients undergoing THA. The WOMAC index was developed specifically for patients with osteoarthritis [12]. It is responsive [3], has high test-retest reliability [13], and has internal consistency [20] in patients undergoing THA. Our study has correlation with the various components of the functional scoring system and established scoring systems such as the WOMAC and SF-36. Our functional score has been validated against well-established scoring systems in patients without hip disease, patients with osteoarthritis of the hip, and patients after THA, confirming its use in a range of patient populations. The functional scoring system is unique in the type of tasks the patients are expected to complete. The tasks measure hip function and are more likely to demonstrate subtle changes than existing scoring systems that look at ability to perform daily activities after hip replacement. No such "tasks" are part of the WOMAC or SF-36 systems. This may explain why the correlations were not all greater than 0.7 in all cases. This fact is perhaps further established by the higher correlation noted between pain components of the functional score and the WOMAC scores. A correlation coefficient, $r > 0.58$, was noted between the

functional scores and the SF-36 scores. This was also true about the correlation between WOMAC and SF-36 scores ($r > 0.53$). Christensen et al. compared a modified WOMAC score (nonarthritic hip score) with the Harris hip score and the SF-12 scoring system. A high correlation was noted with the Harris hip score, whereas only a moderate correlation was noted with the SF-12 score. The moderate correlation of our scoring system compared with the commonly used scoring systems (SF-36 and WOMAC) that were validated against here may reflect the ability of the scoring system to specifically detect the function of the hip. While WOMAC and SF-36 look at daily tasks and well-being, our scoring system specifically looks at high-demand hip function. Conversely, this observation may be explained by a lack of general pain or functional questions in the functional hip score. We believe this feature may be a specific advantage when used to detect hip functional outcome in a high-demand patient.

The functional hip score is easy to use in a clinical setting. The score is concise and takes a maximum of 10 min to complete. It is based on tasks related to musculoskeletal hip function. The tasks are easy to perform in a hospital setup and can form part of the standard physiotherapy perioperative assessment, increasing the compliance of obtaining outcome scores in patients. In this study, the tasks were supervised and scored by a specialist physiotherapist. However, the tasks are easy to teach and learn. They can be easily reproduced by the patient at home. The surgeon or the physiotherapist may thus encourage the patients to fill out the scoring sheets before their regular clinic assessment and keep their own records to monitor improvements in function, especially in the early postoperative period. The functional hip score has high reliability in reproducing the results as noted by the test-retest reliability. This scoring system also has high construct validity to measure a single underlying concept as noted by the values of Cronbach's alpha in our study.

In conclusion, the functional hip score is a valid and reliable score with a high internal consistency. It can be used as a functional tool, especially in active patients undergoing THA. This may be an ideal research tool for a robust assessment of hip function. The functional hip score can detect improvement in function following clinical intervention when scored pre- and postoperatively.

Disclosures Each author certifies that he or she has no commercial associations (e.g., consultancies, stock ownership, equity interest, or patent/licensing arrangements) that might pose a conflict of interest in connection with the submitted article.

One or more of the authors (FSH) has or may receive payments or benefits from a commercial entity (Smith and Nephew, Inc) that may be perceived as a potential conflict of interest.

Each author certifies that his or her institution approved the human protocol for this investigation, that all investigations were conducted in conformity with ethical principles of research, and that informed consent for participation in the study was obtained.

Appendix

Functional Hip score			
NAME:		DATE OF BIRTH:	
HOSPITAL NUMBER		OPERATION:	Right / left
PRE / POST OP TIME	PRE-OP / 6/52 / 3/12 / 6/12 / 1 YEAR / 2 YEARS / 3 YEARS +	TODAY'S DATE:	

1] Single leg stance (1 minute)

- Number of times foot / hand down for balance
 <3 (1 point)
 3-5 (2 points)
 6-8 (3 points)
 >8 (4 points)
- How difficult was that activity on a scale of 1 -10, with 1 being not at all difficult and 10 being impossible to achieve

1 2 3 4 5 6 7 8 9 10
- How painful was that activity on a scale of 1 -10, with 1 being not at all painful and 10 being the most painful

1 2 3 4 5 6 7 8 9 10

2] Timed stair climb x 10 stairs
 ≤3.9 seconds (1 point)
 4-7.9 seconds (2 points)
 8-11.9 seconds (3 points)
 ≤12 seconds (4 points)

- How difficult was that activity on a scale of 1 -10, with 1 being not at all difficult and 10 being impossible to achieve

1 2 3 4 5 6 7 8 9 10
- How painful was that activity on a scale of 1 -10, with 1 being not at all painful and 10 being the most painful

1 2 3 4 5 6 7 8 9 10

3] Lateral step up onto stairs

- Balance assist required
 None ☐
 1 x hand ☐
 2 x hands ☐
 Unable ☐
 Number of steps

- How difficult was that activity on a scale of 1 -10, with 1 being not at all difficult and 10 being impossible to achieve

1 2 3 4 5 6 7 8 9 10

|-----|

- How painful was that activity on a scale of 1 -10, with 1 being not at all painful and 10 being the most painful

1 2 3 4 5 6 7 8 9 10

|-----|

4) Three Forward jumps, standing up between each

Unable to do (4 points)

Stop midway (3 points)

Can do but painful (2 points)

Can do and no pain (1 point)

- How difficult was that activity on a scale of 1 -10, with 1 being not at all difficult and 10 being impossible to achieve

1 2 3 4 5 6 7 8 9 10

|-----|

- How painful was that activity on a scale of 1 -10, with 1 being not at all painful and 10 being the most painful

1 2 3 4 5 6 7 8 9 10

|-----|

5) Three Sideways jumps, standing up between each

Unable to do (4 points)

Stop midway (3 points)

Can do but painful (2 points)

Can do and no pain (1 point)

- How difficult was that activity on a scale of 1 -10, with 1 being not at all difficult and 10 being impossible to achieve

1 2 3 4 5 6 7 8 9 10

|-----|

- How painful was that activity on a scale of 1 -10, with 1 being not at all painful and 10 being the most painful

1 2 3 4 5 6 7 8 9 10

|-----|

References

- Amstutz HC, Thomas BJ, Jinnah R, Kim W, Grogan T, Yale C. Treatment of primary osteoarthritis of the hip. A comparison of total joint and surface replacement arthroplasty. *J Bone Joint Surg Am.* 1984;66:228.
- Ashby E, Grocott MPW, Haddad FS. Outcome measures for orthopaedic interventions on hip. *J Bone Joint Surg Br.* 2008;90:545–549.
- Bachmeier CJ, March LM, Cross MJ, Lapsley HM, Tribe KL, Courtenay BG, Brooks PM; Arthritis Cost and Outcome Project Group. A comparison of outcomes in osteoarthritis patients undergoing total hip and knee replacement surgery. *Osteoarthritis Cartilage.* 2001;9:137–146.
- Bardakos NV, Vasconcelos JC, Villar RN. Early outcome of hip arthroscopy for femoroacetabular impingement: the role of femoral osteoplasty in symptomatic improvement. *J Bone Joint Surg Br.* 2008;90(12):1570–5.
- Beaulé PE, Dorey FJ, Hoke R, Leduff M, Amstutz HC. The value of patient activity level in the outcome of total hip arthroplasty. *J Arthroplasty.* 2006;21:547–552.
- Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol.* 1988;15:1833–1840.
- Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, Westlake L. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ.* 1992;305:160–164.
- Christensen CP, Althausen PL, Mittleman MA, Lee JA, McCarthy JC. The nonarthritic hip score: reliable and validated. *Clin Orthop Relat Res.* 2003;406:75–83.
- Dawson J, Fitzpatrick R, Murray D, Carr A. Comparison of measures to assess outcomes in total hip replacement surgery. *Qual Health Care.* 1996;5:81–88.
- Dawson J, Fitzpatrick R, Frost S, Gundle R, McLardy-Smith P, Murray D. Evidence for the validity of a patient-based instrument for assessment of outcome after revision hip replacement. *J Bone Joint Surg Br.* 2001;83:1125–1129.
- Feeny D, Wu L, Eng K. Comparing Short Form 6D, standard gamble, and Health Utilities Index Mark 2 and Mark 3 utility scores: results from total hip arthroplasty patients. *Qual Life Res.* 2004;13:1659–1670.
- Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. *Med Care.* 1992;30:917–925.
- Kennedy D, Stratford PW, Pagura SMC, et al. Exploring the factorial validity and clinical interpretability of the WOMAC. *Physiotherapy Canada.* 2003;55:160–168.
- Laupacis A, Bourne R, Rorabeck C, Feeny D, Wong C, Tugwell P, Leslie K, Bullas R. The effect of elective total hip replacement on health-related quality of life. *J Bone Joint Surg Am.* 1993;75:1619–1626.
- Mahomed NN, Arndt DC, McGrory BJ, Harris WH. The Harris hip score: comparison of patient self-report with surgeon assessment. *J Arthroplasty.* 2001;16:575–580.
- Martin DP, Engelberg R, Agel J, Swiontkowski MF. Comparison of the Musculoskeletal Function Assessment questionnaire with the Short Form-36, the Western Ontario and McMaster Universities Osteoarthritis Index, and the Sickness Impact Profile health-status measures. *J Bone Joint Surg Am.* 1997;79:1323–1335.
- McHorney CA, Ware JE Jr, Rogers W, Raczek AE, Lu JF. The validity and relative precision of MOS short- and long-form health status scales and Dartmouth COOP charts. Results from the Medical Outcomes Study. *Med Care.* 1992;30(Suppl):MS253–265.
- Ostendorf M, van Stel HF, Buskens E, Schrijvers AJ, Marting LN, Verbout AJ, Dhert WJ. Patient-reported outcome in total hip replacement. A comparison of five instruments of health status. *J Bone Joint Surg Br.* 2004;86:801–808.
- Söderman P, Malchau H. Is the Harris hip score system useful to study the outcome of total hip replacement? *Clin Orthop Relat Res.* 2001;384:189–197.
- Stucki G, Sangha O, Stucki S, Michel BA, Tyndall A, Dick W, Theiler R. Comparison of the WOMAC osteoarthritis index and self-report format of the self administered Lequesne-Algo functional index in patients with knee and hip osteoarthritis. *Osteoarthritis Cartilage.* 1998;6:79–86.
- Ware JE Jr, Sherbourne CD. The MOS 36 item short-form health survey (SF-36). Conceptual framework and item selection. *Med Care.* 1992;30:473–483.
- Ware J Jr, 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.