

# Reliability and Validity of the Cross-Culturally Adapted German Oxford Hip Score

Florian D. Naal MD, Marc Sieverding MD, Franco M. Impellizzeri MS,  
Fabian von Knoch MD, Anne F. Mannion PhD, Michael Leunig MD

Received: 2 March 2008 / Accepted: 1 August 2008 / Published online: 26 August 2008  
© The Association of Bone and Joint Surgeons 2008

**Abstract** There is currently no German version of the Oxford hip score. Therefore we sought to cross-culturally adapt and validate the Oxford hip score for use with German-speaking patients (OHS-D) with osteoarthritis of the hip using a forward-backward translation procedure. We then assessed the new score in 105 consecutive patients (mean age, 63.4 years; 48 women) undergoing THA. We specifically determined: the number of fully completed questionnaires, reliability, concurrent validity by correlation with the WOMAC, Harris hip score, and SF-12, and distribution of floor and ceiling effects. We received 96.6% fully completed questionnaires. An intraclass correlation coefficient of 0.90 and Cronbach's alpha of 0.87 suggested the OHS-D was reliable. Correlation coefficients between the OHS-D and the WOMAC total score, pain subscale, stiffness subscale, and physical function subscale were 0.82, 0.70, 0.68, and 0.82, respectively. OHS-D correlated

with the Harris hip score ( $r = 0.63$ ) and the physical component scale of the SF-12 ( $r = 0.58$ ). We observed no ceiling or floor effects. The OHS-D appeared a reliable and valid measurement tool for assessing pain and disability with German-speaking patients with hip osteoarthritis.

**Level of Evidence:** Level I, diagnostic study. See the Guidelines for Authors for a complete description of levels of evidence.

## Introduction

The traditional approach to outcome assessment after THA has been to measure clinical signs and symptoms. However, this approach fails to reflect the patient's perspective. As such, in recent years, outcome assessment has increasingly focused on patient-reported questionnaires [19]. Such self-report questionnaires should be used to add knowledge and allow more complete assessment of patients' conditions [22]. Self-report questionnaires generally should be short to increase the response rate and decrease the risk of data loss. They also should be reliable, valid, and sensitive to clinical change [19]. The Oxford hip score (OHS), a 12-item, joint-specific, self-administered questionnaire, has been studied extensively since its development and is a reliable, valid, and responsive instrument for assessing hip pain and disability in patients undergoing THA [7, 8, 13, 17, 18, 20].

To avoid population-related and culture-related bias in assessment, when questionnaires developed in one language are to be used in another, it is not sufficient to simply translate the questions. The questionnaires should be adapted cross-culturally to maintain the content and construct validity of the original instrument [21]. Although German is one of the most common European languages

---

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

Each author certifies that his or her institution has 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.

---

F. D. Naal (✉), M. Sieverding, F. von Knoch, M. Leunig  
Department of Orthopedic Surgery, Schulthess Clinic,  
Lengghalde 2, 8008 Zurich, Switzerland  
e-mail: florian.naal@gmail.com

F. M. Impellizzeri, A. F. Mannion  
Department of Research and Development, Schulthess Clinic,  
Zurich, Switzerland

M. Leunig  
University of Berne, Berne, Switzerland

and is spoken by more than 140 million people, there is currently no German version of the OHS.

We therefore created an OHS for use with German-speaking patients with hip osteoarthritis. We chose the OHS because it is short, practicable, reliable, and valid. We specifically asked whether our German version would show similar reliability and concurrent validity with the latter being examined by the strength of the correlation between OHS scores and the scores on other, longer instruments measuring similar constructs.

## Materials and Methods

The cross-cultural adaptation of the OHS was performed according to the guidelines of the American Association of Orthopaedic Surgeons Outcomes Committee [2]. The process involved five stages, each of which was documented with a written report. Step 1 involved forward translation from English to German by an informed translator ([FDN] T1, orthopaedic surgeon, mother tongue German, fluent in English) and an uninformed translator ([MGW] T2, mother tongue German, fluent in English). Step 2 comprised the synthesis of T1 and T2 into one version (T12) with any discrepancies being resolved under the supervision of a methodologist (AFM) who was not involved in the initial translation process. A German language professional verified the accuracy and appropriateness of the language used in the T12 version. In Step 3, two independent backtranslations of the T12 version from German to English were created by native English speakers ([SH] BT1 and [CM] BT2) fluent in German and naive to the outcome measure. Step 4 comprised a consensus meeting of all persons involved in the translation process to resolve any problems, discrepancies, and ambiguities, and to establish the prefinal German version (OHS-D). Step 5 involved pretesting of the German version in 30 consecutive patients (undergoing THA in our hospital) for accuracy of wording and ease of understanding of the questionnaire.

The study involved 105 consecutive German-speaking patients undergoing primary THA in October and November 2007. There were 48 women (46%) and 57 men (54%). The mean age of the patients was  $63.4 \pm 11$  years (range, 33–88 years). There were no differences in the mean age or gender distribution (both  $p > 0.05$ ) between the study sample and our routine patient collective of the last 5 years ( $n = 2500$ ). Our institution is a large orthopaedic hospital with more than 600 primary THAs performed per year. Access to the hospital is open to every patient, and our routine patients are a mixture of urban and rural inhabitants. The study cohort therefore was considered representative. The study was approved by the local ethical committee and all patients provided written informed consent to participate.

We mailed a complete set of questionnaires accompanied by an explanatory letter to the patients 1 week before their admission for surgery. Patients were requested to fill out the questionnaires at home and bring them on the day of admission. After completing the first set, 43 patients volunteered to complete a second questionnaire set for assessment of test-retest reliability. The time between test and retest was approximately 1 week.

Relative reliability concerns the degree to which individuals maintain their position in a sample with repeated measurements [1]. We assessed this type of reliability with the intraclass correlation coefficient (ICC2,1), a two-way random effects model with single measures (absolute agreement) in which variance over the repeated session is considered. Absolute reliability is given by the degree to which repeated measurements vary for individuals (ie, test-to-test noise) [1]. We expressed this type of reliability using the Bland and Altman 95% limits of agreement with the mean difference between duplicate scores representing the bias and the 95% confidence interval representing the random error [4]. Systematic bias was examined using a paired t-test. Heteroscedasticity was examined by plotting the absolute differences between the two sets of scores against their means and calculating the Pearson's correlation coefficient between these two variables; significant correlations indicated the presence of heteroscedasticity [1, 5]. Internal consistency of the German OHS was examined by calculating Cronbach's alpha (CA) [6]. CA indicates the average correlation between all items of a scale and the correlation between each item and the whole scale. The CA can range from 0 (no correlation) to 1 (perfect correlation). We expected CA values greater than 0.8, which were considered good. CA values greater than 0.9 were considered excellent. In the development study, the Bland and Altman's coefficient of reliability was calculated as 7.3 and the CA was 0.84 [7].

The concurrent validity of the translated OHS was examined by analyzing the strength of the correlation between its scores and those of the WOMAC, Harris hip score (HHS), and SF-12 using Spearman's rank correlation coefficients. All scores for the analysis of concurrent validity were completed at administration of the first questionnaire. The OHS is a 12-item instrument with each item scored by the patient on a 1- to 5-point Likert scale [7]. The global score is given by the sum of the scores for all 12 items resulting in values between 12 and 60. The higher the score, the worse the health state. In our study, we recoded the scores into a 0- to 100-point scale with 100 being the best score. The WOMAC is a self-administered, disease-specific measure that contains subscales for pain, stiffness, and physical function [3, 23]. The original global score is calculated as the sum of the scores for each subscale. Scores range from 0 to 20 (pain), 0 to 8 (stiffness),

and 0 to 68 (function). The higher the score, the worse the health state. As for the OHS, the scores were recoded into a 0- to 100-point scale with 100 being the best score. The HHS is a clinician-based, joint-specific assessment tool and requires the surgeon or clinician to grade the patient's pain (44 points), mobility and walking (47 points), range of motion (5 points), and absence of deformities (4 points) [12]. The higher the score, the better the health state. The HHS was recorded once on admission to the hospital. The SF-12 is a self-administered generic measure of quality of life [10, 25]. Scores are transformed into two weighted summary scores for physical function (Physical Component Scale [PCS]) and mental health (Mental Component Scale [MCS]) which can score between 0 and 100 [10, 25]. The higher the score, the better the health state. To examine convergent validity, we hypothesized that the correlation coefficients describing the relationship between the OHS and WOMAC and the HHS and the PCS of the SF-12 would be moderate to high ( $r = 0.50\text{--}0.80$ ). To examine divergent validity, we hypothesized the correlation coefficients describing the relationship between the OHS and the MCS of the SF-12 would be lower than those between the OHS and pain or physical function-related scores and subscales ( $r < 0.50$ ). In their analysis of pre-operative patients the developers reported correlation coefficients between the OHS and the SF-36 domains in the range of  $-0.19$  to  $-0.68$  [7].

The distribution of floor and ceiling effects of the German OHS was determined by calculating the proportion of individuals obtaining the lowest (12) and highest (60) scores, respectively [24]. This indicates the proportion of patients for whom it would not be possible to measure a meaningful improvement (ie, even lower score) or deterioration (ie, even higher score) of their condition, because they are already at the extreme of the range.

Unless otherwise stated, all data are presented as the mean  $\pm$  standard deviation. Normal distribution of the scores was tested using the Shapiro–Wilk W test. Only fully completed questionnaires were used for the analysis; forms with any missing data were excluded. The statistical analysis was performed using the software package SPSS version 13.0 (SPSS Inc, Chicago, IL).

## Results

The forward and backtranslations of the OHS presented no major problems or difficulties with the language. Most discrepancies concerned synonyms for specific expressions, eg, “difficulty  $\rightarrow$  Schwierigkeiten  $\rightarrow$  problems.” Similarly, the phrase “from your hip” was translated into German as “in Ihrer Hüfte” (the verbatim translation “von Ihrer Hüfte” not being appropriate in German), which

**Table 1.** Mean score values and Spearman rank correlation coefficients

Outcome measure	Score values (mean $\pm$ sd)	OHS-D (Spearman R)	p Value
Oxford hip score <sup>†</sup>	48.5 $\pm$ 14.7	—	—
WOMAC total <sup>†</sup>	55.2 $\pm$ 19.9	0.82	< 0.001
WOMAC pain <sup>†</sup>	52.7 $\pm$ 21.5	0.70	< 0.001
WOMAC stiffness <sup>†</sup>	51.8 $\pm$ 24.4	0.68	< 0.001
WOMAC function <sup>†</sup>	53.7 $\pm$ 21.4	0.82	< 0.001
Harris hip score	60.1 $\pm$ 14.8	0.63	< 0.001
SF-12 (PCS)	34.0 $\pm$ 9.1	0.58	< 0.001
SF-12 (MCS)	53.7 $\pm$ 11.2	0.30	0.001

SF-12 = Short Form-12; PCS = physical component scale; MCS = mental component scale; OHS-D = German Oxford hip score; SD = standard deviation. <sup>†</sup> = score values recoded into a 0 to 100 points scale with 100 being the best score.

resulted in “in your hip” being returned in the backtranslation. Pretesting of the German version (OHS-D, Appendix 1) in 30 patients revealed no difficulties in comprehension of the items.

The completion rate of the OKS-D was 96.6%. There was no specific question that consistently was left unanswered. Missing items appeared to arise randomly. Mean scores for the first and second OHS administrations were similar ( $p = 0.83$ ) ( $48.5 \pm 14.7$  versus  $46.4 \pm 15.9$ , respectively). The test-retest reliability was confirmed with an ICC of 0.90 (95% CI, 0.82–0.95). Bland and Altman's limits of agreement suggested no significant bias [ $-2.1$  (95% CI,  $-4.28\text{--}0.01$ );  $p = 0.06$ ] and a random error of  $\pm 13.5$  (total error  $-15.6\text{--}11.4$ ). We observed no heteroscedasticity. Internal consistency was confirmed with a CA of 0.87. Convergent validity for the OHS-D was observed by the moderate to high correlations between OHS-D scores and the other questionnaire scores (Table 1). The strongest correlations were observed between the OHS-D and the WOMAC function score ( $r = 0.82$ ) and the OHS-D and WOMAC total score ( $r = 0.82$ ). The correlation coefficient between the OHS-D and the MCS of the SF-12 was weak ( $r = 0.30$ ), indicating adequate divergent validity. We found no floor or ceiling effects for the OHS-D. Two patients had scores between the lowest value and the random error of measurement ( $0\text{--}13.5$  points), but no patients had scores between the highest value and the random error ( $86.5\text{--}100$  points). The worst score was 6.3 and the best was 85.4, each in one patient.

## Discussion

The traditional approach to outcome assessment after THA has been to measure clinical signs and symptoms which, however, fails to reflect the patient's perspective. Patient

self-report questionnaires should be used to add knowledge and allow more complete assessment of the patients' conditions [22]. The OHS, a 12-item, joint-specific, self-administered questionnaire, has been studied extensively and is a reliable, valid, and responsive instrument for assessing hip pain and disability in patients undergoing THA [7, 8, 13, 17, 18, 20]. Our study (1) cross-culturally adapted and (2) validated the OHS for use with German-speaking patients with hip osteoarthritis.

Before interpreting the results of our study, several limitations must be considered: First, our patient sample represented mainly Swiss German-speaking patients. However, the OHS-D was developed in written German and there are few semantic differences in the use of the written language among the German-speaking countries. Moreover, neither Swiss patients nor German-speaking immigrants had difficulties with wording or understanding of the questionnaire. We therefore do not believe our primarily Swiss-German speaking cohort has introduced a substantial bias. Second, the time between test and retest was relatively short which might have positively biased our reliability results. Finally, this validation was performed in patients with hip osteoarthritis undergoing THA. We believe further investigation of the OHS-D in patients after THA is warranted to concomitantly assess the sensitivity to change of this measure.

Our patients had no major difficulties completing the OHS-D as revealed by detailed interviews of the 30 individuals in the pretest phase and the subsequent high completion rate in the main study of 96.6%. This rate was higher than reported rates [9, 13, 20, 26]. In contrast to the studies of Wood and McLauchlan [26] and McMurray et al. [16], we did not find any specific question that was responsible for noncompletion. In cases with missing data, the entire back page was left unanswered (Questions 7–12). As a consequence, a note was added at the end of the first page that clearly indicates the questionnaire continues on the reverse side of the page.

In accordance with the results reported for the original English version of the OHS [7], the reliability of the OHS-D was high with an ICC of 0.90. The random error of  $\pm 13.5$  we detected was higher than originally reported ( $\pm 7.3$ ) which is explained by the score recoding into a 0- to 100-point scale with 100 being the best score. Using the original scoring method (12 to 60 points with 12 being the best score), the random error was calculated as  $\pm 6.5$ . The random error can be considered the minimal detectable change at the individual level [1]. We found good internal consistency for the OHS-D with a CA of 0.87, similar to the value reported by Dawson et al. (0.84) [7]. The concurrent validity of the OHS-D was confirmed by the strong correlations between its scores and those of the WOMAC pain and function subscales and the WOMAC total score

( $r = 0.70$ – $0.82$ ). This confirms previous findings for the original version of the OHS [11, 18]. In a prospective cohort study on 402 patients (mean age, 61 years), Garbuz et al. reported correlation coefficients of  $r = 0.81$ – $0.87$  between OHS and WOMAC total score, and pain and function subscales [11]. Ostendorf et al. reported correlation coefficients of 0.76 and 0.88 between OHS and WOMAC pain and function subscales in a cohort of 147 patients with a mean age of 68 years [18]. We observed that the correlation coefficient describing the relationship between the OHS-D and the WOMAC stiffness subscale was somewhat lower ( $r = 0.68$ ), which also is consistent with those of Garbuz et al. ( $r = 0.57$ ) [11] and Ostendorf et al. ( $r = 0.63$ ) [18]. We found a moderately high correlation between the scores of the OHS-D and those of the HHS ( $r = 0.63$ ); this was in line with the findings of Kalairajah et al. ( $r = -0.71$ ) who compared the HHS with the OHS in 200 patients (mean age, 68 years) 5 years after THA [13]. The divergent validity of the OHS-D was observed by its low correlation with the mental health domain of the SF-12 (MCS). We observed a coefficient of 0.30, which was slightly lower than the values of Ostendorf et al. ( $r = -0.49$ ) [18] and Garbuz et al. ( $r = -0.49$ ) [11]. The correlation coefficient between the OHS-D and the PCS of the SF-12 in our study ( $r = 0.58$ ) was in line with those of Ostendorf et al., and Garbuz et al. ( $r = -0.53$ ;  $r = -0.60$ ) [11, 18]. The different prefixes for correlation coefficients are explained by the recoding of the scores in our study. Similar to the findings for preoperative patients reported by Garbuz et al. [11], we observed no floor or ceiling effects for the OHS-D.

The mean preoperative OHS scores in our patient sample were notably better than those reported in previous studies, mainly performed in the United Kingdom [7, 8]. The mean preoperative OHS score reported by Dawson et al., was 43.6 [7]; Field et al. reported a mean preoperative value of 41.0 [8] and Ostendorf et al. reported a value of 42.5 for Dutch patients [18]. In our patient sample, when using the original scoring system, the mean preoperative score was only 35.0. This was not the result of age- or gender-related effects because the mean age and gender distribution were comparable among patients in all these studies. One explanation might concern the waiting time for surgery. One study suggests the clinical status may deteriorate while on a waiting list for THA [14]. Ostendorf et al. specified a mean waiting time of 6 months for their patients [18]. In the United Kingdom, where most previous studies using the OHS were done [7, 8], waiting times are approximately 12 to 18 months [14]. In our hospital, in contrast, waiting times for THA typically range from 6 to 12 weeks. Therefore we believe differences in waiting time might contribute to the different preoperative scores for patients from different countries. Geographic and sociocultural differences also



might have contributed to the observed differences; Lingard et al. described different patient expectations and outcomes for patients undergoing surgery in the United States, Australia, and the United Kingdom [15]. However, whether Swiss patients have a better perception of their health state in general is speculative.

Our data show the German version of the OHS (OHS-D) is a practicable, reliable, and valid instrument for self-assessment of pain and function with German-speaking patients with hip osteoarthritis. This study can serve as a model for other non-English speaking investigators for cross-cultural adaptation of outcome measures.

**Acknowledgments** We thank Susan Huber, Charles McCammon, and Moritz Große Wentrup for help with the cross-cultural adaptation process.

### Appendix 1. The German Version of the Oxford Hip Score

#### Oxford Hüfte Score

Bitte beantworten Sie die folgenden 12 Fragen, indem Sie bei jeder Frage die zutreffende Zahl ankreuzen. Wählen Sie nur eine Antwort pro Frage.

Während der letzten 4 Wochen...

1. Wie würden Sie die Schmerzen beschreiben, die Sie üblicherweise in Ihrer Hüfte hatten?

- 1) Keine
- 2) Sehr Gering
- 3) Gering
- 4) Mässig
- 5) Stark

2. Hatten Sie wegen Ihrer Hüfte Schwierigkeiten, sich selbst zu waschen und abzutrocknen (am ganzen Körper)?

- 1) Überhaupt keine Schwierigkeiten
- 2) Sehr geringe Schwierigkeiten
- 3) Mässige Schwierigkeiten
- 4) Extreme Schwierigkeit
- 5) Unmöglich zu tun

3. Hatten Sie wegen Ihrer Hüfte Schwierigkeiten, in ein, bzw. aus einem Auto zu steigen oder öffentliche Verkehrsmittel zu benutzen? (welches Sie eher benutzen)

- 1) Überhaupt keine Schwierigkeiten
- 2) Sehr geringe Schwierigkeiten
- 3) Mässige Schwierigkeiten
- 4) Extreme Schwierigkeit
- 5) Unmöglich zu tun

4. Konnten Sie sich ein Paar Socken, Strümpfe oder Strumpfhosen anziehen?

- 1) Ja, leicht
- 2) Mit geringen Schwierigkeiten
- 3) Mit mässigen Schwierigkeiten
- 4) Mit extremen Schwierigkeiten
- 5) Nein, unmöglich

5. Konnten Sie die Haushaltseinkäufe selbst erledigen?

- 1) Ja, leicht
- 2) Mit geringen Schwierigkeiten
- 3) Mit mässigen Schwierigkeiten
- 4) Mit extremen Schwierigkeiten
- 5) Nein, unmöglich

6. Wie lange konnten Sie gehen, bevor Sie starke Schmerzen in Ihrer Hüfte bekamen (mit oder ohne Stock)?

- 1) Keine Schmerzen /> 30 Minuten
- 2) 16 bis 30 Minuten
- 3) 5 bis 15 Minuten
- 4) Nur zu Hause
- 5) Gar nicht

7. Konnten Sie eine Treppe hinauf gehen?

- 1) Ja, leicht
- 2) Mit geringen Schwierigkeiten
- 3) Mit mässigen Schwierigkeiten
- 4) Mit extremen Schwierigkeiten
- 5) Nein, unmöglich

8. Wie schmerzhaft war es für Sie wegen Ihrer Hüfte, nach einer Mahlzeit wieder vom Tisch aufzustehen?

- 1) Gar nicht schmerzhaft
- 2) Ein wenig schmerzhaft
- 3) Mässig schmerzhaft
- 4) Sehr schmerzhaft
- 5) Unerträglich

9. Haben Sie wegen Ihrer Hüfte beim Gehen gehinkt?

- 1) Selten/nie
- 2) Manchmal oder nur am Anfang
- 3) Oft, nicht nur am Anfang
- 4) Die meiste Zeit
- 5) Die ganze Zeit

10. Hatten Sie plötzliche, starke Schmerzen – „einschiesend“, „stechend“ oder „krampfartig“ – in Ihrer betroffenen Hüfte?

- 1) Nie
- 2) Nur 1 oder 2 Tage
- 3) Einige Tage
- 4) Die meisten Tage
- 5) Jeden Tag

11. Wie sehr haben Schmerzen in Ihrer Hüfte Ihre normale Arbeit (einschliesslich Hausarbeit) beeinträchtigt?
  - 1) Gar nicht
  - 2) Ein wenig
  - 3) Mässig
  - 4) Erheblich
  - 5) Vollständig
12. Wurden Sie nachts im Bett durch Schmerzen in Ihrer Hüfte gestört?
  - 1) Nie
  - 2) Nur 1 oder 2 Nächte
  - 3) Einige Nächte
  - 4) Die meisten Nächte
  - 5) Jede Nacht

## References

1. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* 1998;26:217–238.
2. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine.* 2000;25:3186–3191.
3. 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.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307–310.
5. Bland JM, Altman DG. Measurement error proportional to the mean. *BMJ.* 1996;313:106.
6. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:297–334.
7. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br.* 1996;78:185–190.
8. Field RE, Cronin MD, Singh PJ. The Oxford hip scores for primary and revision hip replacement. *J Bone Joint Surg Br.* 2005;87:618–622.
9. Fitzpatrick R, Morris R, Hajat S, Reeves B, Murray DW, Hannen D, Rigge M, Williams O, Gregg P. The value of short and simple measures to assess outcomes for patients of total hip replacement surgery. *Qual Health Care.* 2000;9:146–150.
10. Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, Bullinger M, Kaasa S, Leplege A, Prieto L, Sullivan M. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol.* 1998;51:1171–1178.
11. Garbuz DS, Xu M, Sayre EC. Patients' outcome after total hip arthroplasty: a comparison between the Western Ontario and McMaster Universities index and the Oxford 12-item hip score. *J Arthroplasty.* 2006;21:998–1004.
12. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am.* 1969;51:737–755.
13. Kalairajah Y, Azurza K, Hulme C, Molloy S, Drabu KJ. Health outcome measures in the evaluation of total hip arthroplasties: a comparison between the Harris hip score and the Oxford hip score. *J Arthroplasty.* 2005;20:1037–1041.
14. Kili S, Wright I, Jones RS. Change in Harris hip score in patients on the waiting list for total hip replacement. *Ann R Coll Surg Engl.* 2003;85:269–271.
15. Lingard EA, Sledge CB, Learmonth ID; Kinemax Outcomes Group. Patient expectations regarding total knee arthroplasty: differences among the United States, United Kingdom, and Australia. *J Bone Joint Surg Am.* 2006;88:1201–1207.
16. McMurray R, Heaton J, Sloper P, Nettleton S. Measurement of patient perceptions of pain and disability in relation to total hip replacement: the place of the Oxford hip score in mixed methods. *Qual Health Care.* 1999;8:228–233.
17. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, Dawson J. The use of the Oxford hip and knee scores. *J Bone Joint Surg Br.* 2007;89:1010–1014.
18. 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.
19. Pynsent PB. Choosing an outcome measure. *J Bone Joint Surg Br.* 2001;83:792–794.
20. Pynsent PB, Adams DJ, Disney SP. The Oxford hip and knee outcome questionnaires for arthroplasty. *J Bone Joint Surg Br.* 2005;87:241–248.
21. Rosemann T, Szecsenyi J. Cultural adaptation and validation of a German version of the Arthritis Impact Measurement Scales (AIMS2). *Osteoarthritis Cartilage.* 2007;15:1128–1133.
22. Stratford PW, Kennedy DM. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *J Clin Epidemiol.* 2006;59:160–167.
23. Stucki G, Meier D, Stucki S, Michel BA, Tyndall AG, Dick W, Theiler R. [Evaluation of a German version of WOMAC (Western Ontario and McMaster Universities) Arthrosis Index][in German]. *Z Rheumatol.* 1996;55:40–49.
24. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34–42.
25. 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.
26. Wood GC, McLauchlan GJ. Outcome assessment in the elderly after total hip arthroplasty. *J Arthroplasty.* 2006;21:398–404.