

## Development of a Valid Simplified Chinese Version of the Oxford Hip Score in Patients With Hip Osteoarthritis

Wei Zheng MD, Jia Li MD, Jinzhu Zhao MD,  
Denghui Liu MD, Weidong Xu MD

Received: 14 August 2013 / Accepted: 19 November 2013 / Published online: 6 December 2013  
© The Association of Bone and Joint Surgeons® 2013

### Abstract

**Background** Although the Oxford Hip Score has been translated and validated in several languages, there is currently no Chinese version of the outcomes measurement. Our study aims to crossculturally adapt and validate the Oxford Hip Score into a simplified Chinese version.

**Questions/purposes** We tested the (1) reliability; (2) validity; and (3) responsiveness of the Chinese version of the Oxford Hip Score.

**Methods** First we translated the Oxford Hip Score into simplified Chinese, then back into English, then held a consensus meeting to achieve the final simplified Chinese version. Then we evaluated the psychometric properties of Chinese version of the Oxford Hip Score in patients undergoing total hip arthroplasty (THA). All patients undergoing THA between July and December 2012 were invited to participate in this study; a total of 108 (79% of

136 invited) did so. To assess the test-retest validity, all participants completed the Chinese version of the Oxford Hip Score again with a 2-week interval. Pearson correlation coefficient was used to evaluate the construct validity between the Chinese version of the Oxford Hip Score and visual analog scale (VAS), Harris hip score, and eight individual domains of the SF-36. Responsiveness was demonstrated by comparing the pre- and postoperative scores of the Chinese version of the Oxford Hip Score.

**Results** The test-retest reliability with intraclass correlation coefficient (0.937) and internal consistency with Cronbach's alpha (0.91) were excellent. The Chinese version of the Oxford Hip Score correlated with the Harris hip score (0.89,  $p < 0.01$ ), VAS ( $-0.79$ ,  $p < 0.01$ ), and Physical Functioning (0.79,  $p < 0.01$ ) and Bodily Pain (0.70,  $p < 0.01$ ) domains of SF-36, which suggested construct validity. No floor or ceiling effects were found. The effect size and standardized response mean values were 3.52 and 3.31, respectively, indicating good responsiveness.

**Conclusions** The Chinese version of the Oxford Hip Score showed good reliability, validity, and responsiveness in evaluating standard Chinese-speaking patients with hip osteoarthritis undergoing THA. It can be used by clinical surgeons as a complement to the traditional outcome measures.

---

Each author certifies that he or she, or a member of his or her immediate family, has no funding or 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.

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research* editors and board members are on file with the publication and can be viewed on request.

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.

Wei Zheng and Jia Li contributed equally to this work as cofirst authors.

---

W. Zheng, J. Li, J. Zhao, D. Liu, W. Xu (✉)  
Department of Orthopedics, Changhai Hospital, Second Military  
Medical University, 168 Changhai Road, Shanghai 200433,  
China  
e-mail: xuwshanghai@gmail.com

### Introduction

Hip osteoarthritis (OA) is common, painful, and sometimes disabling [16]. To determine the influence of the disease and its treatments on pain, function, and quality of life in patients with hip OA, surgeons increasingly use patient-reported questionnaires. These should be reliable, valid, and sensitive to clinical changes [23]. The Oxford Hip

Score (OHS) is a 12-item, hip-specific, self-reported questionnaire for patients with hip diseases. It has been widely used as an outcome measure of functional ability, daily activities, and pain from the patient's perspective [5]. There are 12 domains, and each is scored using a self-reported 5-point Likert scale; the OHS' sum score therefore ranges from 0 (worst) to 48 (best). The OHS has been studied extensively and has proven to be reliable, valid, and responsive for patients [5, 21]. It also has been translated and validated in several languages like German, Dutch, Japanese, French, and Italian [7, 10, 17, 20, 24].

When one reliable, valid questionnaire is being used in populations with different cultures, it is necessary to test the psychometric properties of the questionnaire rather than simply translating the content. China has the largest population (approximately 1.3 billion), and Chinese is one of the most common languages in the world; however, there has not been a Chinese version of the OHS (OHS-C) so far.

Therefore, we aimed to perform an intercultural adaptation of OHS for the Chinese-speaking population with hip OA and evaluated the psychometric properties of the Chinese version in Chinese patients undergoing THA. Specifically, we tested the (1) reliability; (2) validity; and (3) responsiveness of the Chinese version of the OHS.

## Materials and Methods

### Translation and Crosscultural Adaptation

The translation of the original OHS followed previous published guidelines [1, 11]. The process was formed in five steps: Step 1—forward translation. The forward translation from English to simplified Chinese was performed independently by three bilingual translators who were native Chinese. Two of the translators were orthopaedic surgeons in our hospital (authors of the article, WZ and JL); another one was a professional bilingual translator with no medical background unaware of the study purpose. Step 2—synthesis of the translation. The first Chinese version of the OHS was obtained after a consensus meeting of the three translators. Step 3—backtranslation. Three native English speakers (YJ, FA, GD) with a medical background, fluent in Chinese, blind to the previous English version of OHS, translated the Chinese version of OHS back into English. Step 4—a consensus meeting with all translators was held to compare the backtranslation with the first Chinese version, original English version, and to resolve discrepancies, ambiguities, or any other problems to reach a prefinal Chinese version of the OHS. Step 5—test the prefinal version of OHS on 30 consecutive patients with hip OA to see if there were any problems with the prefinal version. All the translators should discuss the problem and

**Table 1.** Characteristics of participants

Characteristics	Total (N = 108)
Age (years, mean $\pm$ SD)	66 $\pm$ 11
Range	35–87
Sex, number (%)	
Female	63 (58%)
Male	45 (42%)
Hip osteoarthritis duration (years, mean $\pm$ SD)	5.8 $\pm$ 2.5
Range	1–12

develop the final Chinese version of the OHS (OHS-C) and perform further psychometric testing.

### Psychometric Assessments and Statistical Analysis

#### Participants

Between July and December 2012, all 136 standard Chinese-speaking patients undergoing THA were invited to participate in this study. The inclusion criteria were as follows: age > 18 years, able to read and speak Chinese, primary hip OA diagnosed based on the criteria of the American College of Rheumatology, and willing to receive a THA in our hospital. Patients were excluded if they were unable or unwilling to complete the questionnaire or if they had symptomatic OA in the other lower limb joints, a history of lower limb or spine surgery, inflammatory arthritis, spondyloarthritis, or severe lung, heart, or other diseases. Finally, a total of 108 (79% of those invited, 63 women and 45 men) who met the prespecified inclusion criteria did so. Mean age of participants was 66 years (range, 35–87 years). Duration of hip OA was 5.8 $\pm$ 2.5 years (range, 1–12 years) (Table 1). The population was in accordance with the article proposed by Terwee et al. [22] that the study should enroll at least 100 patients for internal consistency analysis and 50 patients for floor or ceiling effects, reliability, and validity analysis. All the 108 patients signed informed consent to participate in the study and the clinical research ethics committee of our hospital approved the study.

#### Instruments

The OHS is widely used to assess patients with diseases of the hip and it includes 12 items (each scores on a 0–4 Likert scale). The questionnaire generates an overall score ranging from 0 to 48 with a higher score representing better hip status. The OHS has been translated and validated into several languages [7, 10, 17, 20, 24].

To determine construct validity, we compared the OHS with the Harris hip score, the SF-36, and the visual analog scale (VAS) score for pain. The Harris hip score (HHS), a joint-specific health status questionnaire, is frequently used by clinicians to assess the outcome of the hip. The HHS contains four domains: pain, function, deformity, and ROM ranging from 0 (maximum disability) to 100 (no disability) [12]. The SF-36 is a general health status-measuring questionnaire that contains eight domains: Physical Functioning, Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional, and Mental Health. The SF-36 has been translated and validated in Chinese populations in many studies. Each subscale ranges from 0 to 100 and higher scores represent better health status [15, 25, 27]. The VAS is a simple and widely used method to measure patients' intensity of pain. It allows patients to rate pain intensity along a 100-mm line ranging from "no pain" (at the left end) to "worst pain" (at the right end) [8].

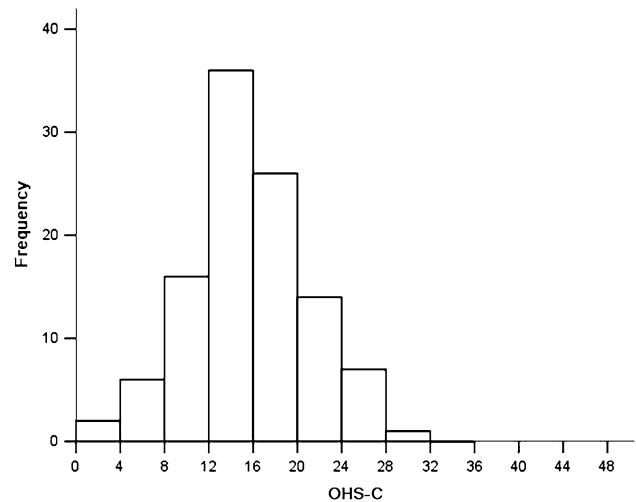
Participants completed the OHS-C, HHS, VAS, and SF-36 in an outpatient room of orthopaedics in our hospital. Two weeks later, when they were in the hospital waiting for surgery, they were asked to complete the questionnaires for the second time. Six months after the surgeries, the participants were required to complete the OHS-C for the third time.

*Acceptability and Score Distribution*

To evaluate acceptability, all the patients were asked if there were any difficulties filling in the questionnaire. The data were checked for missing or multiple responses. The completeness of the OHS-C and the time needed to complete the OHS-C were also measured. The average time required to complete the OHS-C was  $96 \pm 24$  seconds. All participants completed the OHS-C and there were no missing responses or difficulties observed. Scores of OHS-C ranged from 3 to 31 (Fig. 1). We also summed the scores of the VAS, HHS, and SF-36 (Table 2).

*Reliability*

The reliability property was assessed by internal consistency and test-retest reliability. Internal consistency was measured by Cronbach's alpha. Cronbach's alpha > 0.7 is considered good reliability [22]. We measured test-retest reliability by comparing scores of the first and second time. The health status of patients with such a chronic disease is unlikely to change too much during 2 weeks without medical intervention. People also would not recall the answers they chose before. The intraclass correlation



**Fig. 1** This is the distribution of the OHS-C scores.

**Table 2.** Score distribution of the OHS-C, HHS, VAS, and SF-36 (N = 108)

Instrument	Minimum	Maximum	Mean	SD
OHS-C	3	31	15	5
HHS	10	67	37	11
VAS (mm)	20	100	77	20
SF-36				
PF	0	55	28	14
RP	0	75	6	16
BP	0	62	35	11
GH	15	67	43	8
VT	30	95	67	12
SF	13	88	40	14
RE	0	100	63	16
MH	16	100	68	13

OHS-C = Chinese version of the Oxford Hip Score; HHS = Harris hip score; VAS = visual analog scale; PF = Physical Functioning; RP = Role-Physical; BP = Bodily Pain; GH = General Health; VT = Vitality; SF = Social Functioning; RE = Role-Emotional; MH = Mental Health.

coefficient (ICC) was used to assess the test-retest reliability, where a value > 0.8 is considered of good reproducibility [9]. Bland-Altman, describing the mean scores of the two assessments and differences between them, was also used to assess whether there was systematic bias between the test and retest of the OHS-C [2, 3].

*Validity*

Construct validity was evaluated by calculating the Pearson correlation coefficients among the OHS-C and HHS, VAS, and eight domains of the SF-36. The correlations were

judged as poor ( $r = 0-0.20$ ), fair ( $r = 0.21-0.40$ ), moderate ( $r = 0.41-0.60$ ), good ( $r = 0.61-0.80$ ), or excellent ( $r = 0.81-1.0$ ). Now that the OHS was interculturally adapted to evaluate the physical health of the hip, we hypothesized that the OHS-C correlated strongly with the physical health-related domains (Physical Functioning, Bodily Pain) of the SF-36 and weakly with the mental health-related domains (Vitality, Role-Mental Health, Role-Emotional) of the SF-36. Floor and ceiling effects were also considered significant if  $> 15\%$  of all the participants achieved the lowest (0) or highest (48) possible score on the OHS-C [18].

### Responsiveness

The responsiveness [4, 13, 19] of the OHS-C was obtained by comparing the preoperative scores and 6-month postoperative scores. We calculated the effect size by using the SD of preoperative OHS-C scores divided by the mean change between preoperative scores and postoperative scores [18]. We also calculated the standardized response mean by using the SD of the changes between pre- and postoperative divided by mean of the changes.

SPSS Version 13.0 (SPSS Inc, Chicago, IL, USA) was used to analyze the datum of all the questionnaires.

## Results

### Reliability

#### Internal Consistency

The internal consistency was good. The Cronbach's alpha was 0.91 for the overall OHS-C and ranged from 0.90 to 0.91 if an item was deleted. The item total correlation ranged from 0.43 to 0.77, which also indicated good correlation between each item and the overall OHS-C (Table 3).

#### Test-retest

The OHS-C showed excellent test-retest reliability. Mean score of the retest was  $15.7 \pm 5.0$ , which was similar to the first test ( $15.3 \pm 5.3$ ;  $p > 0.05$ ). ICC for the test-retest was 0.937 (95% confidence interval, 0.909–0.957; Table 4). Bland-Altman plot (Fig. 2) showed no systematic bias. The limits of agreement ranged from  $-4.01$  to  $3.20$ . It also indicated good reproductivity of the OHS-C [3].

**Table 3.** Correlation of each item and total OHS-C scores

Question	Mean $\pm$ SD	Item total correlation	Cronbach's alpha if item deleted
Describe the pain in the hip	$0.70 \pm 0.57$	0.70	0.90
Troubled by pain in bed at night	$1.54 \pm 0.72$	0.61	0.91
Sudden, severe pain	$1.65 \pm 0.75$	0.65	0.90
Limp when walking	$0.56 \pm 0.57$	0.43	0.91
Distance of walk	$1.37 \pm 0.71$	0.77	0.90
Climb a flight of stairs	$1.03 \pm 0.56$	0.76	0.90
Put on socks, stocking or tights	$1.48 \pm 0.65$	0.69	0.90
Pain when stand up from a chair	$1.42 \pm 0.60$	0.56	0.91
getting in and out of a car	$1.61 \pm 0.56$	0.56	0.91
Washing and drying yourself	$1.92 \pm 0.76$	0.65	0.90
Household shopping by own	$0.98 \pm 0.56$	0.70	0.90
Interfere with the usual life	$1.06 \pm 0.45$	0.76	0.90
Cronbach's alpha: 0.91			

OHS-C = Chinese version of the Oxford Hip Score.

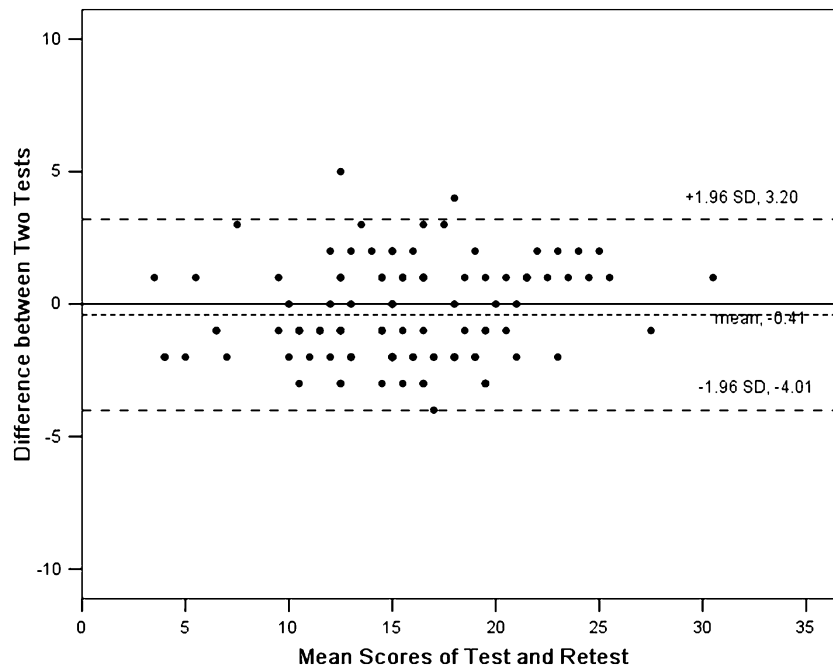
**Table 4.** Intraclass correlation coefficient between the test and retest groups (n = 108)

Instrument	Mean score (SD)		ICC (95% CI)
	First test	Second test	
OHS-C	15.31 (5.34)	15.71 (5.04)	0.937 (0.909–0.957)

OHS-C = Chinese version of the Oxford Hip Score; ICC = intraclass correlation coefficient; CI = confidence interval.

### Validity

The result demonstrated that the correlation between OHS-C and HHS ( $0.89$ ,  $p < 0.01$ ) was excellent. The OHS-C also correlated well with the VAS ( $-0.79$ ,  $p < 0.01$ ) and the Physical Functioning ( $0.79$ ,  $p < 0.01$ ) and Bodily Pain ( $0.70$ ,  $p < 0.01$ ) domains of the SF-36. These data indicated convergent validity. A correlation between OHS-C and Role-Physical ( $0.52$ ,  $p < 0.01$ ), General Health ( $0.55$ ,  $p < 0.01$ ), and Social Functioning ( $0.51$ ,  $p < 0.01$ ) domains of the SF-36 was moderate. However, the correlation between the OHS-C and Vitality ( $0.31$ ,  $p < 0.01$ ), Role-Emotional ( $0.31$ ,  $p < 0.01$ ), and Mental Health ( $0.29$ ,  $p < 0.01$ ) domains of the SF-36 was weak, indicating divergent validity. We also observed that the OHS-C showed a better correlation with SF-36 than HHS (Table 5).



**Fig. 2** These are Bland-Altman plots of test-retest reliability of the OHS-C. The interval of two assessments was 2 weeks. Dashed lines show the 95% (mean ± SD) limits of agreement.

**Table 5.** Pearson correlations among the OHS-C, HHS, VAS, and SF-36 (n = 108)

Instrument	OHS-C	HHS
HHS	0.89*	
VAS	-0.79*	-0.75*
SF-36		
PF	0.79*	0.69*
RP	0.52*	0.43*
BP	0.70*	0.59*
GH	0.55*	0.49*
VT	0.31*	0.35*
SF	0.51*	0.48*
RE	0.31*	0.32*
MH	0.29*	0.30*

\* Correlation is significant at the 0.01 level (two-tailed); OHS-C = Chinese version of the Oxford Hip Score; HHS = Harris hip score; VAS = visual analog scale; PF = Physical Functioning; RP = Role-Physical; BP = Bodily Pain; GH = General Health; VT = Vitality; SF = Social Functioning; RE = Role-Emotional; MH = Mental Health.

**Responsiveness**

The Chinese version of the OHS showed good responsiveness to treatment. The responsiveness of the OHS-C was evaluated by comparison of the pre- and postoperative scores of the THA group. The mean score of OHS-C improved from 15 ± 5 to 34 ± 4 (p < 0.01). The mean of changes was 19 ± 5. The effect size and standardized response mean for OHS-C were 3.52 and 3.31, respectively.

**Discussion**

In China, clinical surgeons are paying more attention to self-reported outcome assessment. Several hip-specific instruments have been translated and crossculturally adapted into Chinese, including the Hip Disability and Osteoarthritis Outcome Score [26]. At present, there is no agreement for which questionnaire should be used to evaluate the status of patients with hip OA. The OHS is widely used as a joint-specific measure for patients with hip OA [14], but to our knowledge, this widely used tool has not been validated in a Chinese population. The purpose of this study therefore was to interculturally adapt the OHS into Chinese and to evaluate the psychometric properties of the OHS-C in a Chinese population with hip OA undergoing THA. We found the Chinese version of the OHS to be a valid tool, demonstrating a high degree of reliability, validity, and responsiveness.

Before discussing our results further, there are some limitations of our study that should be considered. First, the participants did not represent the entire Chinese population with hip OA. Most of the patients recruited had severe hip OA and intended to undergo THA. However, there was enough variability in the population to demonstrate responsiveness, and no floor or ceiling effects were observed. Second, we translated the OHS into a standard simplified Chinese language, the official language of China, but traditional Chinese language was also widely



used in several southern areas in China. So it is necessary to translate and validate the OHS into traditional Chinese language in the future. Third, all of the participants underwent THA. We did not assess the responsiveness in patients receiving conservative treatments. Thus, more validation research in patients with hip OA with other treatments would be required.

The Cronbach's alpha correlation coefficient for the OHS-C (0.914) indicated excellent internal consistency, which was equivalent to other studies of OHS [6, 7, 10, 17, 20, 24]. The Pearson coefficients of item total (ranging from 0.427 to 0.770) also indicated good correlation between item and overall score. As for the test-retest reliability, ICC for the OHS-C (0.937; 95% confidence interval, 0.909–0.957) and Bland-Altman plot (Fig. 2) was considered of good reproducibility. It was in accordance with other validation studies [10, 17, 20].

Construct validity was demonstrated by calculating the correlation between OHS-C scores and HHS, VAS, and eight individual domains of SF-36 scores. The OHS-C correlated significantly with HHS (0.890) and VAS (−0.788), which suggested the OHS-C measured similar aspects to HHS and VAS. We also observed that OHS-C showed a significant correlation with Physical Functioning (0.79,  $p < 0.01$ ) and Bodily Pain (0.70,  $p < 0.01$ ) domains of the SF-36 and a weak correlation with Vitality (0.31,  $p < 0.01$ ), Role-Emotional (0.31,  $p < 0.01$ ), and Mental Health (0.29,  $p < 0.01$ ) domains of the SF-36 (Table 5). The result of construct validity was consistent with previous validation studies [6, 7, 10, 17, 24]. No floor or ceiling effects were observed in the pre- and postoperative patients, similar to previous studies [10, 17].

The responsiveness, or sensitivity to clinical change, is the most important characteristic in prospective outcome study. The result showed that the OHS-C was able to detect change after surgical treatment with excellent responsiveness. The effect size of the OHS-C was 3.52. Compared with those who received hyaluronic injection (effect size 1.98), patients who received a THA showed a better effect size of the OHS [17]. It was also better than the effect size of patients receiving a THA in other studies of OHS [6, 10]. Our explanation was that the participants in our study were in worse health status than those of other validation studies, which might lead to better responses to surgical treatment.

In summary, we found that the OHS could be interculturally adapted into Chinese with good psychometric properties. As a self-reported questionnaire, the Chinese version of the OHS is a joint-specific, reliable, valid instrument for a Chinese population with hip OA undergoing THA. Therefore, we suggest that the OHS-C can be used by surgeons in practice to evaluate the impact of hip OA and its treatments on patients' pain and function.

**Acknowledgments** We thank the staff from our outpatient clinics and the patients participating in the study. We also thank Yang Jiao, Francis Aaron, and Gregory Dole for help with the translation process.

## References

1. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine Phila Pa* 1976. 2000;25:3186–3191.
2. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.
3. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:135–160.
4. Davies GM, Watson DJ, Bellamy N. Comparison of the responsiveness and relative effect size of the Western Ontario and McMaster Universities Osteoarthritis Index and the Short-Form Medical Outcomes Study Survey in a randomized, clinical trial of osteoarthritis patients. *Arthritis Care Res*. 1999;12:172–179.
5. 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.
6. Dawson J, Fitzpatrick R, Murray D, Carr A. Comparison of measures to assess outcomes in total hip replacement surgery. *Quality in Health Care*. 1996;5:81–88.
7. Delaunay C, Epinette JA, Dawson J, Murray D, Jolles BM. Cross-cultural adaptations of the Oxford-12 HIP score to the French speaking population. *Orthop Traumatol Surg Res*. 2009;95:89–99.
8. de Nies F, Fidler MW. Visual analog scale for the assessment of total hip arthroplasty. *J Arthroplasty*. 1997;12:416–419.
9. Fleiss JL, Shrout PE. The effects of measurement errors on some multivariate procedures. *Am J Public Health*. 1977;67:1188–1191.
10. Gosens T, Hoefnagels NH, de Vet RC, Dhert WJ, van Langelaan EJ, Bulstra SK, Geesink RG. The 'Oxford Hip Score': the translation and validation of a questionnaire into Dutch to evaluate the results of total hip arthroplasty. *Acta Orthop*. 2005;76:204–211.
11. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46:1417–1432.
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. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol*. 2000;53:459–468.
14. 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.
15. Li L, Wang HM, Shen Y. Chinese SF-36 Health Survey: translation, cultural adaptation, validation, and normalisation. *J Epidemiol Community Health*. 2003;57:259–263.
16. Loureiro A, Mills PM, Barrett RS. Muscle weakness in hip osteoarthritis: a systematic review. *Arthritis Care Res (Hoboken)*. 2013;65:340–352.
17. Martinelli N, Longo UG, Marinozzi A, Franceschetti E, Costa V, Denaro V. Cross-cultural adaptation and validation with reliability, validity, and responsiveness of the Italian version of the Oxford Hip Score in patients with hip osteoarthritis. *Qual Life Res*. 2011;20:923–929.
18. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4:293–307.

19. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19:539–549.
20. Naal FD, Sieverding M, Impellizzeri FM, von Knoch F, Mannion AF, Leunig M. Reliability and validity of the cross-culturally adapted German Oxford Hip Score. *Clin Orthop Relat Res.* 2009;467:952–957.
21. 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.
22. 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.
23. Thorborg K, Roos EM, Bartels EM, Petersen J, Hölmich P. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review. *Br J Sports Med.* 2010;44:1186–1196.
24. Uesugi Y, Makimoto K, Fujita K, Nishii T, Sakai T, Sugano N. Validity and responsiveness of the Oxford Hip Score in a prospective study with Japanese total hip arthroplasty patients. *J Orthop Sci.* 2009;14:35–39.
25. Wang W, Lopez V, Ying CS, Thompson DR. The psychometric properties of the Chinese version of the SF-36 health survey in patients with myocardial infarction in mainland China. *Qual Life Res.* 2006;15:1525–1531.
26. Wei X, Wang Z, Yang C, Wu B, Liu X, Yi H, Chen Z, Wang F, Bai Y, Li J, Zhu X, Li M. Development of a simplified Chinese version of the Hip Disability and Osteoarthritis Outcome Score (HOOS): cross-cultural adaptation and psychometric evaluation. *Osteoarthritis Cartilage.* 2012;20:1563–1567.
27. Zhou Z, Yang L, Chen Z, Chen X, Guo Y, Wang X, Dong X, Wang T, Zhang L, Qiu Z, Yang R. Health-related quality of life measured by the Short Form 36 in immune thrombocytopenic purpura: a cross-sectional survey in China. *Eur J Haematol.* 2007;78:518–523.