Osteoarthritis and Cartilage 20 (2012) 1563–1567

Osteoarthritis and Cartilage

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Brief report

Development of a simplified Chinese version of the hip disability and osteoarthritis outcome score (HOOS): cross-cultural adaptation and psychometric evaluation

X. Wei [a], Z. Wang [a], C. Yang [a], B. Wu [a], X. Liu, H. Yi, Z. Chen, F. Wang, Y. Bai, J. Li, X. Zhu, M. Li*

Orthopaedic Department of Changhai Hospital, Second Military Medical University, Shanghai 200433, People's Republic of China

ARTICLE INFO

Article history:
Received 19 January 2012
Accepted 22 August 2012

Keywords:
HOOS
Cross-cultural adaptation
Psychometric property

Osteoarthritis (OA) has a profound impact on health-related quality of life[1]. Increasing importance has been attached to utilization of disease-specific, self-reported outcome measures[2], such as the Hip Disability and Osteoarthritis Outcome Score (HOOS) instrument[3]. China is the most populous country in the world with 1.3 billion people. Hence, we translated and adapted the HOOS into a Simplified Chinese version (SC-HOOS) and validated it in a cohort of native Chinese-speaking patients with hip OA, relative to the Short Form Health Survey (SF-36), a visual analog scale (VAS), and the Harris hip score (HHS) test. Psychometric testing for internal consistency, test–retest reliability, construct validity, and responsiveness was conducted. The SC-HOOS showed satisfactory internal consistency, test–retest reliability, construct validity, and responsiveness when evaluated in Chinese-speaking patients with hip OA.

Participants and data analysis

A total of 131 consecutive patients (58 men, 73 women) with a diagnosis of primary hip OA were recruited from the Department of Orthopedics of our medical university between December 2010 and August 2011, and enrolled in accordance with the quality criteria described by Terwee *et al.*[4] They had a mean age of 51.3 (9.1) years and a mean body mass index (BMI) of 25.2 (3.0) kg/m^2. The project was approved by the Human Research Ethics Committee of our institution, and each patient signed a written informed consent. The inclusion criteria were: age >18 years, ability to read and speak Chinese, and primary hip OA diagnosis according to the criteria of the American College of Rheumatology[5]. The exclusion criteria were: history of leg or spine surgery, tumors, infection, rheumatologic disease, ankylosing spondylitis, and/or neuropathologies; inability or unwillingness to complete questionnaires independently. Patients in whom surgical treatment [total hip replacement (THR) subgroup] was deemed necessary ($N = 52$) were allowed 3 months postoperatively to finish their questionnaires.

Statistical analysis was performed using the Statistical Package for the Social Sciences, version 18.0 (SPSS, Chicago, IL). Mean values are reported with standard deviations (SDs). Intraclass correlation coefficient (ICC) values are reported with 95% confidence intervals (CIs). P values of <0.05 were considered significant.

Translation and cross-cultural adaptation

The translation and cross-cultural adaptation were performed according to previously published guidelines regarding dual forward translation, synthesis of the dual translations using resolution by consensus, backward translation into English to reveal any discrepancies, reconciliation by expert committee consensus, and a test of the pre-final SC-HOOS[6,7]. Our examination of the final SC-HOOS (see Appendix) was consistent with recommendations for cross-cultural validation studies of patient-reported outcomes[8].

* Address correspondence and reprint requests to: M. Li, Orthopaedic Department of Changhai Hospital, Second Military Medical University, No 168, Changhai Road, Shanghai 200433, People's Republic of China. Tel: 86-21-81873387 (O); Fax: 86-21-81873398.
E-mail address: limingch@21cn.com (M. Li).
[a] These authors contributed equally this paper.

1063-4584/$ – see front matter
http://dx.doi.org/10.1016/j.joca.2012.08.018

Psychometric assessments

All 131 patients completed the SC-HOOS, the SF-36, VAS, and HHS in an outpatient hospital room. Patients were asked to finish the SC-HOOS first, before the other tests, and the time frame for completing the SC-HOOS was 15 min. The HOOS includes five subscales: pain (10 items), other symptoms (10 items), function in daily living (ADL) (17 items), function in sports and recreation (Sport/Rec) (four items), and hip-related quality of life (QoL) (four items). A five-point Likert scale was used and each item was given a score of 0 (none), 1 (mild), 2 (moderate), 3 (severe), or 4 (extreme). All of the item scores within each subscale were summed, divided by the maximum score, and then deducted from 100, such that more extreme symptoms resulted in a larger deduction from 100, and thus a lesser subscale score. The normalized scores, from 0 (indicating extreme symptoms) to 100 (indicating no symptoms), for each subscale were plotted in an outcome profile for each participant.

The HHS is a multidimensional observational assessment that contains questions about pain, function, deformity, and range of motion, with a total score ranging from 100 (no disability) to 0 (maximum disability)[9]. The SF-36 contains eight domains: physical functioning (PF), role-physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role-emotional (RE), and mental health (MH). Each raw subscale scores was transformed to a 100-point scale. The SF-36 has been translated into Chinese and thoroughly tested. Finally, the VAS allows patients to rate pain intensity along a 100-mm line ranging from “no pain” to “worst pain imaginable”.

Score distribution, acceptability, and internal consistency

Floor and ceiling effects exceeding 15% were considered significant[4]. The SC-HOOS subscale scores were well-distributed, with no floor or ceiling effects.

To evaluate acceptability, patients were asked about any difficulties that had been encountered. The data were checked for missing or multiple responses and the completeness of the SC-HOOS was calculated. Missing data were treated as recommended by Nilsdotter *et al.*[10] Most (127/131; 96.9%) of the patients responded that they did not have any difficulties with completing the SC-HOOS. The remaining four patients left the question unanswered. The average time for patients to finish the SC-HOOS was 10.4 (3.2) min, similar to that reported previously. The total numbers of improperly answered items in the total test group, the retest subgroup, and the THR subgroup were relatively few at 352/5240 (3.6%), 143/2400 (6.0%), and 97/2080 (4.7%), respectively. The correct completion rates for the entire SC-HOOS were 96.4%, 94.0%, and 95.3% for total test group, the retest subgroup, and the THR subgroup, respectively.

Internal consistency of the SC-HOOS subscales was evaluated by calculating Cronbach’s α coefficient, where $\alpha > 0.80$ and $\alpha > 0.90$ were regarded as good and excellent, respectively[4]. The Cronbach’s α coefficients for the subscales (see Table I) were high (0.865–0.968), especially for the pain and ADL subscales, indicating good internal consistency.

Test–retest reliability

Sixty patients were randomly selected to be in the retest subgroup according to a computer generated randomized number table. These patients were asked to complete the SC-HOOS again at home 7 days after they had completed it the first time, and then to return it by mail once they finished it. A 1-week interval was chosen because it is too brief for obvious post-treatment clinical changes to be apparent, and also because 1 week is the time that THR patients needed to wait for surgery, enabling them to participate in the reliability evaluation. An ICC (two-way random effects model) was calculated to quantify test–retest reliability. An ICC

Table I
Internal consistency, construct validity, reliability, and responsiveness of the SC-HOOS

Parameter	SC-HOOS subscale (No. items)				
	Symptoms (5)	Pain (10)	ADL (17)	Sport/Rec (4)	QoL (4)
Internal consistency					
Cronbach's α	0.883	0.939	0.968	0.865	0.87
Construct validity indicated by Pearson correlation coefficient, r (P value), vs indicated instruments					
SF-36 domains					
PF	0.771 (<0.0001)	0.728 (<0.0001)	0.769 (<0.0001)	0.733 (<0.0001)	0.734 (<0.0001)
RP	0.420 (<0.0001)	0.441 (<0.0001)	0.451 (<0.0001)	0.418 (<0.0001)	0.427 (<0.0001)
BP	0.636 (<0.0001)	0.701 (<0.0001)	0.628 (<0.0001)	0.666 (<0.0001)	0.712 (<0.0001)
GH	0.508 (<0.0001)	0.485 (<0.0001)	0.510 (<0.0001)	0.412 (<0.0001)	0.446 (<0.0001)
VT	0.230 (0.0082)	0.291 (0.0008)	0.276 (0.0014)	0.211 (0.0157)	0.261 (0.0026)
SF	0.463 (<0.0001)	0.458 (0.0001)	0.473 (<0.0001)	0.429 (0.0001)	0.479 (<0.0001)
RE	0.248 (0.0043)	0.252 (0.0037)	0.266 (0.0021)	0.239 (0.0061)	0.308 (0.0003)
MH	0.221 (0.0113)	0.182 (0.0374)	0.190 (0.0299)	0.264 (0.0023)	0.217 (0.0129)
VAS	−0.765 (<0.0001)	−0.786 (<0.0001)	−0.766 (<0.0001)	−0.714 (0.0001)	−0.777 (<0.0001)
HHS	0.898 (<0.0001)	0.848 (<0.0001)	0.887 (<0.0001)	0.827 (0.0001)	0.893 (<0.0001)
Test–retest reliability, mean (SD) or ICC value (CI range)					
Test score	46.3 (18.0)	46.6 (17.2)	46.2 (18.1)	42.5 (16.9)	44.6 (19.7)
Retest score	48.2 (16.9)	47.3 (16.8)	45.8 (17.6)	44.1 (17.0)	42.4 (19.8)
Score change	1.8 (6.0)	0.8 (4.0)	−0.4 (5.4)	1.6 (8.9)	−2.2 (8.2)
ICC (95% CI)	0.940 (0.902–0.964)	0.973 (0.955–0.984)	0.956 (0.921–0.973)	0.862 (0.780–0.915)	0.913 (0.859–0.947)
Responsiveness pre-THR vs 3 months after THR, mean (SD)[*]					
Pre-THR score	27.8 (11.5)	31.8 (8.7)	30.0 (10.6)	27.7 (10.6)	27.7 (9.8)
Post-THR score	57.4 (9.1)	60.8 (11.1)	58.3 (9.6)	54.5 (9.6)	53.7 (8.4)
Change	29.6 (9.5)	29.0 (12.8)	28.3 (13.1)	26.8 (12.3)	29.4 (13.4)
ES	2.57	3.33	2.67	2.53	3.01
SRM	3.12	2.27	2.16	2.18	2.19

[*] $N = 52$; higher scores represent less pain. For comparison, the HHS yielded values of 28.7 (6.9), 51.3 (15.5), 22.6 (9.6), 3.28, and 2.35, respectively, for the responsiveness parameters.