Journal of Research and Practice on the Musculoskeletal System JRPMS

Original Article

Translation, adaptation and validation of the 12-item Oxford hip score into Arabic

Aliaa Khaja[1], Owayed Almutairi[2], Abdulaziz Alkudair[1], Awdhah Alsamhan[1]

[1]Division of Orthopedic Trauma, Department of Orthopedic Surgery, Al-Razi Orthopedic Hospital, Kuwait City, Kuwait;
[2]Division of Orthopedic Trauma, Department of Orthopedic Surgery, Farwaniya Hospital, Kuwait City, Kuwait

Abstract

Objectives: To determine the ease of using the Arabic translation of the score while rendering the Arabic version valid for clinical use by Arab patients. **Methods**: 110 patients participated. The internal consistency tests were performed using Cronbach's alpha. Test-retest reliability (intra-correlation coefficient), convergent construct validity, convergent validity, floor & ceiling effects, and responsiveness was also calculated. In order to measure the level of agreement, Bland-Altman Plot, forest Plots were performed. **Results**: To test the instrument reliability, Cronbach's alpha was obtained. It proved excellent for the three testing occasions – α_1=0.97, α_2=0.97, and α_3=0.97. Intra-class correlation coefficient was fair with a score of 0.58 (95% CI 0.38-0.73). Factor analysis was performed to test for factor validity - principal axis factoring, with no rotation. As hypothesized, the factor analysis yielded one factor, which explains 78% of the variance. Floor effect was recorded for 1% of the patients, and showed a ceiling effect of 8%, 2%, and 3% for the first, second, and third weeks of testing respectively. **Conclusion**: Arabic version of OHS could potentially be used as diagnostic tool for patients with hip problems regarding information about the overall condition of the patient at a given time.

Keywords: Arabic version, Cross-cultural, KOOS, Osteoarthritis outcome score, Oxford hip score

Introduction

Osteoarthritis is a disease of the joint that affects the adult population. Clinical experience has shown that it is the most common joint disease. The incidence and prevalence of osteoarthritis increases with age globally[1-3]. Osteoarthritis has a great impact on national economies due to the high cost of treatment and surgical techniques and the quality of life of the patients'[4]. According to the World Health Organization (WHO) Scientific Group on Rheumatic Disease, it is estimated that 10 percent of the global population who are above 60 years have significant medical disorders that can be linked to osteoarthritis[5]. Because there is an increase in prevalence and incidence with age, life expectancy and obesity will result in an increase in OA incidence[6-9]. Studies have proven the age and sex-standard incidence of hip osteoarthritis to be 88 per 100.000 person-years[2].

The epidemiology of osteoarthritis is multifactorial, encompassing biological, biomechanical and genetic factors[10]. The prevalence is higher in women over 60 years of age than it is in men of the same age group[11,12]. Medical professionals have encountered challenges determining the pain intensity and degenerative progression of the disease[13]. Management of osteoarthritis include lifestyle modification, and pharmacological therapy. This pinpoint the crucial nature of quantifying the degree of pain[14,15].

Patients with hip osteoarthritis experience pain and also complain about functional impairment during daily activities such as climbing the stairs and walking[16,17]. Strength deficits, stiffness, disturbances in gait, loss of joint movement, and decreased walking velocity are common clinical findings that are peculiar to hip OA[8]. Defects in physical function and pain do have a negative effect on the patient's quality of life[16,18].

The authors have no conflict of interest.

Corresponding author: *Aliaa Khaja, MD, Division of Orthopedic Trauma, Department of Orthopedic Surgery, Al-Razi Orthopedic Hospital, Kuwait City, Kuwait*

E-mail: *aliaa.khaja@googlemail.com*

Edited by: *Konstantinos Stathopoulos*

Accepted *21 April 2020*

There is a gradual increase in the usage of patient-reported questionnaires for determining the effects that the disease and its treatment has on the patients' quality of life. The questionnaires must be valid, sensitive to medical changes, and also reliable[19].

With Oxford hip score (OHS), the clinician can determine the specific forms of pain and mobility problems encountered by the patient with hi problems[6,20]. The Oxford hip score is a hip specific, 12-item, self-reported questionnaire specially designed for patients that are eligible for total hip replacement (THR). Oxford hip score has been widely applied as an outcome measure of daily activities, pain, and functional ability from the hip OA patient's perspective[21]. Extensive research has been carried out on the OHS and results from these researches proves that it is reliable and valid, as well as responsive in patients with hip OA[6,21-23]. The 1996 version (which was the original) received an update in 2007 with the introduction of a new scoring system[6,7]. The Oxford Hip Score has been translated in several languages including German, Dutch, Japanese, French, Korean, Italian, and Chinese[24-31]. This particular study aims to investigate the validity of the translated version of the OHS into Arabic. This would prove beneficial to orthopaedic healthcare providers as well as patients where Arabic is the primary language. It is a chance for healthcare establishments to incorporate the OHS tool in their services.

Methods and materials

Translation

The translation process was in concordance with the literature guidelines suggested by (Guillemin[32] et al. 1993, Mathias[33] et al. 1994 Wild[34] 2005 et al and Epstein[35] 2015 et al). The translation panel consisted of two bilingual orthopaedic surgeons, an independent translation agency and Arabic proof reader. The OHS was first translated into Arabic, then back translated, and revised by the translation panel once more. A pilot study was then conducted on 10 bilingual patients at the clinic. The patients were chosen at random to fill out the questionnaire. Then we investigated for the patients' interpretation of each item, language ease, and understanding of the concepts and assessed for need of assistance when filling out the questionnaire before proceeding to launch our full-scale investigation. Permission from the original author of the Oxford Hip score was obtained and were involved in all the stages of this process.

Participants

Data was gathered randomly from 110 patients who visited the arthroplasty clinics in Al-Razi Orthopedic Hospital over a period of 3 weeks. They had completed the Oxford-12 Item Knee Score questionnaire. Patients gave their written consent to use the information they had provided for the purpose of this study. The average age of the participants is 44.3 years, with standard deviation of 15.4 years; implying that majority of the sample is between 30 and 60 years of age. The youngest participant was 16, and the oldest was 76 years of age. The patient was given two patient related outcome questionnaires adapted for Arabic speakers. The Kuwaiti Ministry of Health Ethical Committee is the main authority of patient record keeping, and they had approved this study.

Questionnaires

Oxford-12 Item Hip Score

This questionnaire consists of 12 questions, all of which are focused on the problems patients are experiencing in their hip. To answer these questions repondents use five point Likert-type scale, where 1 indicates that the respondent is not experiencing any difficulties, and 5 indicates that the patient experiences the highest level of difficulty. One score is calculated for each patient by summing the responses to all of the 12 items. Therefore, a score of 60 means that the patient is experiencing severe hip pain.

All 110 patients have completed Oxford-12 Item Hip Score in at least two different occasions prior to surgical intervention (T1 and T2), and 109 of them completed it the third time, after undergoing a total hip replacement (T3). There was a 2 week interval between the first two occasions they had filled the questionnaires, and over a month interval for the 3rd-time test takers.

Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[36]

There are 24 Likert-type questions here, and unlike Oxford-12, each patient has scores from three different subscales. First subscale – pain – has 5 questions (score range 5-20), 2 questions address stiffness (score range 2-8), and physical function has 17 questions (range 17-68). A minimum score on each of the subscales means that patient essentially has not felt any discomfort in his/her hip (if any), and a higher score suggests greater disability. This is the only valid Arabic Patient related outcome score available to our knowledge. The survey was taken in two different occasions, with a time period of 2 weeks between them.

Patient Burden and feasibility

We recorded the average time it took each participant to fill out the questionnaire as well as if they required any assistance during the process as part of the patient-burden investigation.

The feasibility was determined by missing or incomplete responses and these entries were not included in the study.

Data analysis and psychometric scale properties

All the analyses were conducted using IBM SPSS Statistics 21 for Windows.

The test-retest reliability of the questionnaire was examined by using Cronbach's alpha, and interclass correlation coefficient. The patients completed the

		N[1]	Min[2]	Max[3]	Mean	SD[4]	Sk[5]	Ku[6]	Floor effect	Ceiling effect
OHS	Week 1	110	12	60	38.18	12.125	0.157	-0.758	1%	8%
	Week 2	110	12	60	31.63	9.969	0.829	0.685	2%	0%
	Week 3	109	12	60	29.82	10.164	0.828	0.685	1%	3%

Note: [1]Sample size; [2]Minimum; [3]Maximum; [4]Standard deviation; [5]Skewness; [6]Kurtosis.

Table 1. Descriptive statistics of Oxford Hip Score questionnaire.

Subscales		Scores						Change*	ICC (95% CI)	Cronbach's alpha (95% CI)
		First assessment		Second assessment		Third assessment				
		Mean	SD	Mean	SD	Mean	SD			
WOMAC	Pain	53.22	15.90	63.17	18.85			9.95	0.581 (0.234 - 0.760)	0.735 (0.379 - 0.864)
	Stiffness	53.38	16.87	63.55	18.50			10.17	0.593 (0.230 - 0.772)	0.745 (0.375 - 0.872)
	Phyiscal Function	53.31	16.39	62.91	18.60			9.60	0.623 (0.262 - 0.793)	0.768 (0.416 - 0.884)
Oxford Hip Score		38.18	12.13	31.63	9.97	29.82	10.16	-8.36	0.583 (0.357 - 0.730)	0.808 (0.625 - 0.890)
OHS (Rescaled 0-100)		54.54	25.26	36.72	20.77	37.12	21.17	-17.42		

* Minus sign in OHS means that the condition of patient has been improved over time (lower score = improvement) / Plus sign in WOMAC means that the condition of patient has been worsend over time (higher score = Deterioration).

Table 2. Mean, standard Deviation, Change, ICC between different assessments of each subscale.

questionnaire on three different occassions, so Cronbach's alpha was calculated for all three situations to determine internal validity.

It is hypothesized that Oxford-12 evaluates hip disability as a unidimensional constuct. This was investigated with use of factor analysis.

Convergent and divergent construct validity were tested by using the spearman's correlation coeffiecnts for the 12-items of the oxford hip score and the WOMAC. According to Naal el al.[37]. German version of OHS showed the highest correlation with physical function and then Pain subscales of WOMAC (r>0.80).

Content validity was tested by examining the shape of data distribution, as well as floor and ceiling effects. Floor effect represents the percentage of patients who had the lowest score (12), and ceiling effect is the percentage of those with the highest (60).

Correlation with WOMAC subscales were estimated to evaluate convergent validity of the Oxford-12 Item. If it was a valid instrument to measure hip pain, its scores should have positive correlation with all the subscales from the questionnaire.

Results

Translation

Two final versions were produced by the translation panel were significantly similar and were merged into one. The back-translation of final version was stable as well. There were no cultural discrepancies between the original and the translated version of the questionnaire.

WOMAC questionnaire

Before we proceed with the results from the Oxford-12 questionnaire, it is fair to note the psychometric properties of the instrument that is already being used in Arabic speaking countries for hip pain assessment.

Reliability - calculated using Cronbach's alpha - was 0.98 for the pain subscale, 0.98 for the stiffness, and 0.99 for the physical function subscale. These are the values obtained for the first testing. For the second testing, reliability was 0.99, 0.97, and 0.99 (pain, stiffness, and physical function, respectively).

In order to check content validity, floor and ceiling effects were examined. 10% of the patients have recorded floor effect on pain subscale, 14% on stiffness subscale, and

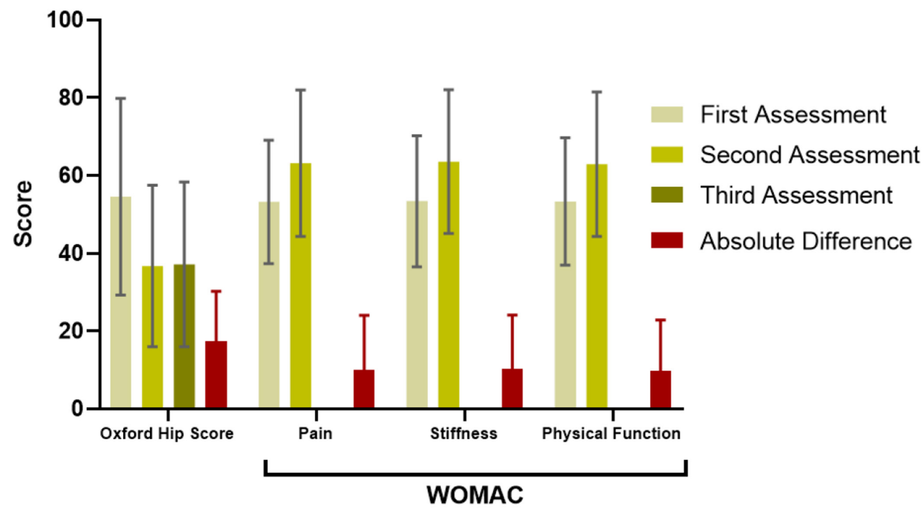

Figure 1. Mean score and absolute difference along with their standard deviations during 3 different assessments for OHS and 2 different assessments for WOMAC questionnaire. Decrease of mean score in OHS= improved condition & Increase of mean score in WOMAC= worsened condition.

12% on the physical function. On the other hand, 3% have recorded ceiling effect on pain subscale, 3% on stiffness subscale, and 3% on the physical function.

Oxford-12 Item Hip Score questionnaire

To test the reliability of the instrument, Cronbach's alpha was obtained. For each of the three testing occasions it was excellent – α_1=0.97, α_2=0.97, and α_3=0.97. Intra-class correlation coefficient was fair with the score of 0.58 (95% CI 0.38-0.73), but it is worth noting that there were two and a half weeks between successive testing situations, so it could reflect changes in patients' clinical condition.

In order to test factor validity, factor analysis was performed - principal axis factoring, with no rotation. As hypothesized, the factor analysis yielded one factor, which explained 78% of the variance.

Floor effect was recorded for 1% of the patients, and 8% showed a ceiling effect in the first week of testing. Two and a half weeks later, 2% of them showed ceiling effect, and there was no floor effect recorded. On the third testing, 1% recorded floor effect, and 3% of the patients had the highest possible score (ceiling effect).

A 2 weeks' test-retest reliability of OHS was applied to the present manuscript. Of the 110 patients that fulfilled the questionnaire, 110 responded to the second assessment after the initial evaluation, and 109 patients responded to the last assessment.

Test-retest reliability was also performed using Intra-class Correlation (ICC). The results (Table 2) indicated that OHS has a moderate intra-class correlation with 0.583 (95% CI 0.357, 0.730). Considering the value of 0.808 (95% CI 0.625 – 0.890) for Cronbach's alpha, the internal consistency of the three assessments were proven to be high.

	WOMAC		
	Pain	Stiffness	Physical function
Week 1			
Oxford-12 Item Hip Score	.48**	.48**	.48**
Week 2			
Oxford-12 Item Hip Score	.45**	.47**	.53**
*Note: ** Correlation is significant at the 0.01 level (2-tailed).*			

Table 3. Convergent validity of the Oxford-12 Item Hip Score questionnaire (Spearman's rank correlation coefficient).

In order to be able to compare the results of WOMAC questionnaire with those from OHS, it was important to standardize the scores of WOMAC to the range of 0-100. In addition, OHS score which were in the range of 12-60, were rescaled to 0-100 to match the WOMAC scores. Figure 1 illustrates the change and the mean level of different subscales during different assessments which were conducted 2 weeks apart from each other. It is visually evident that the mean score of OHS decreased which is related to less pain and symptoms. At the same time the WOMAC mean score is showing an upward trend, which is related with more pain and in general worsened conditions of the patient. This illustrates a visual disagreement between

QUESTIONNAIRES		OXFORD HIP SCORE (OHS)			TOTAL
		Stable	Deterioration	Improvement	
WOMAC	Stable	2.80%	7.48%	27.10%	37.38%
	Improvement	0.00%	4.67%	8.41%	13.08%
	Deterioration	1.87%	5.61%	42.06%	49.53%
TOTAL		4.67%	17.76%	77.57%	100.00%

Table 4. Responsiveness and agreement between the two questionnaires.

Questionnaire	Subscales	Effect Size (Cohen's d)	95% CI*		SRM	95% CI*	
WOMAC	Pain	0.571	0.387	0.751	0.406	0.358	0.434
	Stiffness	0.574	0.395	0.749	0.411	0.366	0.436
	Physical Function	0.547	0.378	0.709	0.410	0.363	0.434
OHS		0.747	0.568	0.962	0.423	0.391	0.445

** Bootstrap confidence interval (1000 iterations; random number seed: 978).*

Table 5. Effect Sizes and SRMs for the WOMAC subscales and OHS. Bars represent the 95% confidence intervals.

the two questionnaires when it comes to describing the change of condition.

As can be seen in the table below, there is a medium to large positive correlations between Oxford-12 Item Score on one side, and all the subscales from the WOMAC questionnaire on the other. This shows that patients with high scores on WOMAC have high scores on Oxford-12.

Responsiveness

14 patients (13.1%) reported overall relevant improvement in their condition by responding to the WOMAC questionnaire, while 53 patients (49.5%) reported worsening of their condition, and 40 of participants remained stable (37.4%).

On the other hand, only 5 patients (4.67%) reported to remain stable by responding to OHS questionnaire. The majority of them (77.57%%) believed their condition to be improved, and only 17.76% of the them reported relevant deterioration after 2 weeks. In addition, it is relevant to note that 50 patients (46.7%) showed contradictory results (45 patients improved according to OHS and worsened according to WOMAC, while 5 patients showed the opposite). 8 patients (7.48%) believed that their condition was worsened according to OHS, while according to WOMAC their conditions were not changed (Table 3).

Effects are often used to give meaning to change over time in terms of 'trivial' (ES<0.20), 'small' (ES≥0.20 <0.50),'moderate' (ES≥0.50 <0.80) or 'large' (ES≥0.80) change. Cohen introduced this 'matched pairs' effect size, which was later renamed the standardised response mean (SRM) by Liang et al.[38]. According to responsiveness test, WOMAC subscales show similar responsiveness (SRM=0.41) between first and second measurement. In comparison to WOMAC, OHS showed better responsiveness with SRM=0.423. This is important to note, however, that responsive change of both questionnaires are very similar and the differences are not considerable.

Level of agreement between WOMAC & HHS

One of the best methods to measure the level of agreement between two measurement methods is Bland-Altman plot. In this method, the mean difference between WOMAC and HHS are plotted as a function of mean of WOMAC and OHS. As shown in the graphs, overall mean difference between WOMAC and OHS shows that there is a systemic bias between two questionnaires (M=15.7, 95% CI12.93, 18.47, p<0.001). In order to test this result, linear regression was performed with mean difference between WOMAC and OHS as a dependent variable and mean value of WOMAC and OHS as independent variable. The result of linear regression also indicates statistically significant difference between the two measurement methods (β= -0.24, 95% CI -0.403 – -0.079, t=-2.95, p=0.004).

First and last measurement of both methods are also compared together with the help of Bland-Altman plot, to investigate whether there will be any change over time to the systemic bias between the two methods. The results indicate that in the first measurement there is a systemic

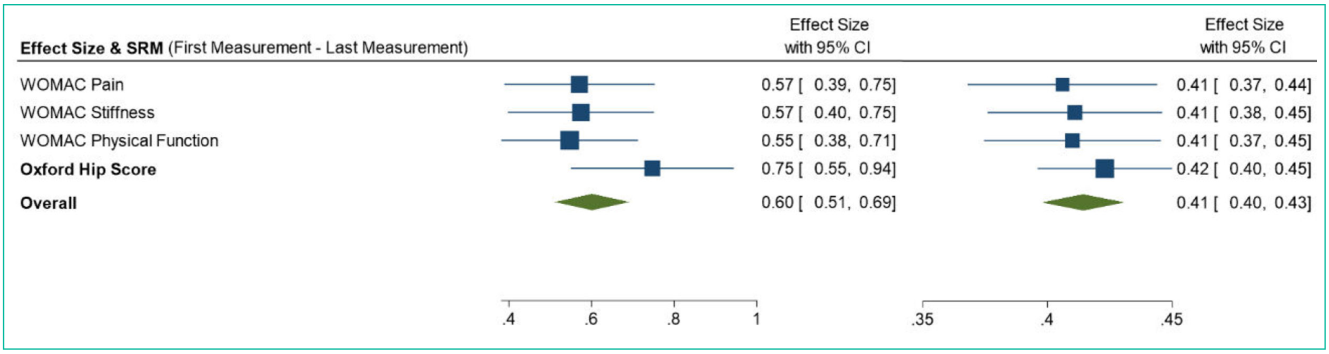

Figure 2. Forest Plot of Effect Sizes and SRMs for the WOMAC subscales and OHS. Bars represent the 95% confidence intervals.

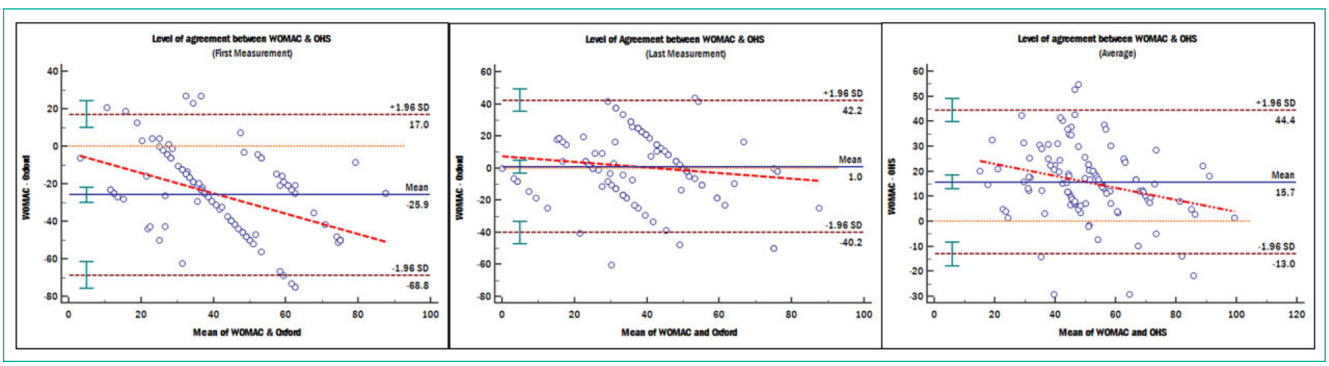

Figure 3. Bland-Altman Plot to demonstrate the level of agreement between OHS and WOMAC (First, last and average assessments). Linear regression line is also drawn to better demonstrate the systemic bias between the two methods

bias between the two methods (M=-25.9, 95% CI -30.01, -21.74, p<0.001), the performed linear regression also confirms this bias (β=-0.54, 95% CI -0.74 – -0.33, t=-5.27, p<0.001). It means that OHS increasingly overestimates the worsened conditions in comparison to WOMAC. However, in the last measurement, the slope of the regression line decreases and became statistically insignificant (β=-0.17, 95% CI -0.417 – 0.006, t=-1.423 p=0.157).

Discussion

The process of translating a score is far from simple, it is a length process that follows specific guidelines, recommendations from the original authors, and accounting for cultural differences. Our survey had an excellent response rate for both the test and retest (95%).

The primary objective of this study was to create a reliable and valid Arabic version of OHS by translation and adaptation. For this purpose, the Arabic version of OHS is compared to the efficacy and results of WOMAC questionnaire. Preliminary validity and reliability tests revealed that there is moderate correlation between WOMAC subscales and OHS, which indicated that they are related in the right direction. It means lower score is associated with less pain and symptoms in both WOMAC and OHS.

However, according to Altman and Bland's views regarding the correct analysis of the data gathered in studies of this type, it is not enough to use the correlation coefficient between the two measurements as a measure of agreement[38]. They pointed out that methods can correlate well yet disagree greatly, as would occur if one method read consistently higher than the other. For this reason, Bland-Altman Plot was used to measure the level of agreement between WOMAC and OHS. The Bland-Altman plots indicated that there is systemic bias between WOMAC and OHS. And the linear regression illustrated that with increasing mean score, Arabic version of OHS tends to underestimate the results of WOMAC. According to Tugay et al.[39], support the use of the Turkish version of OHS as a reliable and valid outcome instrument in Turkish-speaking patients with osteoarthritis of the hip. The illustrated forest plots, and effect sizes, showed that OHS score was generally higher than WOMAC scores. In general, the results of both methods lead the surgeon to the right direction when it comes to information about the overall condition of the patient at a given time, since there is a moderate correlation between both methods, however, it is important to note that there is a potential probability that when the pain is increasing, OHS tends to overestimate the level of the deterioration in comparison to WOMAC. It is also important to be cautious using OHS when the change direction of patient's condition is investigated, since there is a potential probability that the false signal about the improvement or deterioration of condition is reported in comparison to WOMAC.

Conclusion

The primary purpose of this study was to create a reliable and valid Arabic version of OHS by translation and adaptation. For this purpose, the Arabic version of OHS is compared to the efficacy and results of WOMAC questionnaire. Its reliability - calculated both through Cronbach's alpha and ICC - was good.

Correlation with WOMAC subscales are medium to large, which points to its convergent validity. Bland-Altman plot indicated that there is a systemic bias between OHS and WOMAC, which was significantly decreased over time, so that there has been no systemic bias in the last measurement. Overall, Arabic version of OHS could potentially be used as diagnostic tool for patients with hip problems, when it comes to information about the overall condition of the patient at a given time, however, it is important to be cautious using OHS when the follow-up and the recovery process of patient's condition is investigated, since firstly, there is a potential probability that the level of deterioration of the patient's condition will be overestimated by OHS, and secondly, around 40% of patients who were believed to have worsened condition according to OHS, were improving according to WOMAC.

References

1. Michael JW, Schluter-Brust KU, Eysel P (2010). The epidemiology, etiology, diagnosis, and treatment of osteoarthritis of the knee. Dtsch Arztebl Int 107(9):152–162.
2. Oliveria SA, Felson DT, Reed JI, Cirillo PA, Walker AM (1995). Incidence of symptomatic hand, hip, and knee osteoarthritis among patients in a health maintenance organization. Arthritis Rheum 38(8):1134–1141.
3. Felson DT, Zhang Y (1998) An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. Arthritis Rheum 41(8):1343–1355.
4. Dunlop DD, Manheim LM, Song J, Chang RW (2001) Arthritis prevalence and activity limitations in older adults. Arthritis Rheum 44(1):212–221.
5. Woolf AD, Pfleger B (2003) Burden of major musculoskeletal conditions. Bull World Health Organ 81(9):646–656.
6. Dawson J, Fitzpatrick R, Carr A, Murray D (1996) Questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg Br 78B(2):185–190.
7. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, Dawson J (2007) The use of the Oxford hip and knee scores. J Bone Joint Surg Br 89B (8):1010–1014.
8. Bijlsma JW, Berenbaum F, Lafeber FP (2011) Osteoarthritis: an update with relevance for clinical practice. Lancet 377(9783):2115–2126.
9. Dawson J, Linsell L, Zondervan K, Rose P, Randall T, Carr A, Fitzpatrick R (2004) Epidemiology of hip and knee pain and its impact on overall health status in older adults. Rheumatology 43(4):497–504.
10. Veronese N, Cereda E, Maggi S, Luchini C, Solmi M, Smith T, Stubbs B (2016). Osteoarthritis and mortality: A prospective cohort study and systematic review with meta-analysis. Seminars in Arthritis and Rheumatism 46(2):160–167.
11. Glyn-Jones S, Palmer AJR, Agricola R, Price AJ, Vincent TL, Weinans H, Carr AJ (2015). Osteoarthritis. The Lancet 386(9991):376–387.
12. Zheng H, Chen C (2015). Body mass index and risk of knee osteoarthritis: systematic review and meta-analysis of prospective studies. BMJ Open 5(12), e007568.
13. de Rooij M, van der Leeden M, Heymans MW, Holla JF, Häkkinen A, Lems WF, Roorda LD, Veenhof C, Sanchez-Ramirez DC, de Vet HC, Dekker J (2016). Course and predictors of pain and physical functioning in patients with hip osteoarthritis: Systematic review and meta-analysis. Journal of Rehabilitation Medicine 48(3), 245–252.
14. Runhaar J, Rozendaal RM, van Middelkoop M, Bijlsma HJW, Doherty M, Dziedzic KS, Bierma Zeinstra S. (2017). Subgroup analyses of the effectiveness of oral glucosamine for knee and hip osteoarthritis: a systematic review and individual patient data meta-analysis from the OA trial bank. Annals of the Rheumatic Diseases 76(11), 1862–1869.
15. French HP, Smart KM, Doyle F (2017). Prevalence of neuropathic pain in knee or hip osteoarthritis: A systematic review and meta-analysis. Seminars in Arthritis and Rheumatism 47(1), 1–8.
16. Dagenais S, Garbedian S, Wai EK (2009) Systematic review of the prevalence of radiographic primary hip osteoarthritis. Clin Orthop Relat Res 467(3):623–637.
17. Krauss I, Steinhilber B, Haupt G, Miller R, Grau S, Janssen P (2011) Efficacy of conservative treatment regimens for hip osteoarthritis - evaluation of the therapeutic exercise regime "Hip School": a protocol for a randomised, controlled trial. BMC Musculoskelet Disord 12:270.
18. Pereira D, Peleteiro B, Araujo J, Branco J, Santos RA, Ramos E (2011). The effect of osteoarthritis definition on prevalence and incidence estimates: a systematic review. Osteoarthr Cartil OARS Osteoarthr Res Soc 19(11):1270–1285.
19. Bieler T, Magnusson SP, Kjaer M, Beyer N (2014) Intra-rater reliability and agreement of muscle strength, power and functional performance measures in patients with hip osteoarthritis. J Rehabil Med 46(10):997–1005.
20. Fitzpatrick R, Morris R, Hajat S, Reeves B, Murray DW, Hannen D, Rigge M, Williams O, Gregg P (2000) The value of short and simple measures to assess outcomes for patients of total hip replacement surgery. Qual Health Care 9(3):146–150.
21. Thorborg K, Roos EM, Bartels EM, Petersen J, Ho ¨lmich P (2010) Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review. Br J Sports Med 44:1186–1196.
22. Dawson J, Fitzpatrick R, Frost S, Gundle R, McLardy-Smith P, Murray D (2001) Evidence for the validity of a patient-based instrument for assessment of outcome after revision hip replacement. J Bone Joint Surg Br 83(8):1125–1129.
23. Dawson J, Fitzpatrick R, Murray D, Carr A (1996) Comparison of measures to assess outcomes in total hip replacement surgery. Qual Health Care QHC 5(2):81–88.
24. Delaunay C, Epinette JA, Dawson J, Murray D, Jolles BM (2009) Cross-cultural adaptations of the Oxford-12 HIP score to the French

speaking population. Orthop Traumatol Surg Res OTSR 95(2):89–99.

25. Gosens T, Hoefnagels NH, de Vet RC, Dhert WJ, van Langelaan EJ, Bulstra SK, Geesink RG (2005) The "Oxford Heup Score" The translation and validation of a questionnaire into Dutch to evaluate the results of total hip arthroplasty. Acta Orthop 76(2):204–211.
26. Lee YK, Chung CY, Park MS, Lee KM, Lee DJ, Lee SC, Koo KH (2012) Transcultural adaptation and testing of psychometric properties of the Korean version of the Oxford hip score. J Orthop Sci Off J Jpn Orthop Assoc 17(4):377–381.
27. Martinelli N, Longo UG, Marinozzi A, Franceschetti E, Costa V, Denaro V (2011) 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 Int J Qual Life Asp Treat Care Rehabil 20(6):923–929.
28. Naal FD, Sieverding M, Impellizzeri FM, von Knoch F, Mannion AF, Leunig M (2009) Reliability and validity of the cross-culturally adapted German Oxford hip score. Clin Orthop Relat Res 467(4):952–957.
29. Ostendorf M, van Stel HF, Buskens E, Schrijvers AJ, Marting LN, Verbout AJ, Dhert WJ (2004) Patient-reported outcome in total hip replacement. A comparison of five instruments of health status. J Bone Joint Surg Br 86(6):801–808.
30. Uesugi Y, Makimoto K, Fujita K, Nishii T, Sakai T, Sugano N (2009) Validity and responsiveness of the Oxford hip score in a prospective study with Japanese total hip arthroplasty patients. J Orthop Sci Off J Jpn Orthop Assoc 14(1):35–39.
31. Zheng W, Li J, Zhao J, Liu D, Xu W (2014) Development of a valid simplified Chinese version of the Oxford hip score in patients with hip osteoarthritis. Clin Orthop Relat Res 472(5):1545–1551.
32. Guillemin F, Bombardier C, Beaton D (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. Journal of Clinical Epidemiology 46(12), 1417–1432.
33. Mathias SD, Fifer SK, Patrick DL (1994). Rapid translation of quality of life measures for international clinical trials: avoiding errors in the minimalist approach. Quality of Life Research 3(6), 403–412.
34. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P (2005). Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. Value in Health 8(2), 94–104.
35. Epstein J, Santo RM, Guillemin F (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. Journal of Clinical Epidemiology 68(4), 435–441.
36. Klassbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score. An extension of the Western Ontario and McMaster Universities Osteoarthritis Index Scand J Rheumatol 32 (2003), pp. 46-51.
37. Naal FD, Sieverding M, Impellizzeri FM, von Knoch F, Mannion AF, Leunig M (2009). Reliability and validity of the cross-culturally adapted German Oxford hip score. Clinical orthopaedics and related research, 467(4), 952–957.
38. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Medical Care 1990;28(7):632-42
39. Tuğay BU, Tuğay N, Güney H, et al. Arch Orthop Trauma Surg (2015) 135: 879.