Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Harris Hip Score

ALIAA Khaja
alrazi hospital
Owayed AL-Mutairi
Farwaniya Hospital
abdulaziz Alkhudair
al razi hospital
Awdhah Abdulkarim (a.alsamhan099@hotmail.com)
alrazi hospital

Research article

Keywords: Harris Hip Score, pain and function levels, hip pathologies

Posted Date: February 12th, 2020

DOI: https://doi.org/10.21203/rs.2.23408/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Background The Harris Hip Score (HHS) is a widely used Patient-Related Outcomes score. It measures pain and function levels in patients with hip pathologies. Objectives The main objective of this study is to translate and culturally adapt the HHS into Arabic, and to further assess the reliability and validity of translated version Material & Methods 110 patients participated in this survey. 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 are performed. Results Test reliability for the first testing situation - 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. For the second testing, reliability was 0.99, 0.97, and 0.99 (pain, stiffness, and physical function, respectively). This only proves that WOMAC is an instrument with good reliability. Same calculation of Cronbach's alpha was essential to testing the reliability of the Harris Hip Score. For each of the three testing occasions the reliability was very good or excellent – α 1 = 0.92, α 2 = 0.91, and α 3 = 0.90. Intra-class correlation coefficient was good with the score of 0.76 (95% CI 0.44-0.88). Conclusion Overall, Arabic version of HOOS could be used as diagnostic tool for patients with hip problems, when it comes to information about the overall condition of the patient, especially about the improvement or deterioration, however, it is important to be cautious using HHS when the change magnitude of patient's condition is investigated, since there is a potential probability that the level of improvement of the patient's condition will be overestimated by HHS.

Background

Patient-Related Outcomes (PROs) have emerged as useful tools for measuring medical conditions, has have been proven to be extremely useful in musculoskeletal disease clinics.[1] These well-structured questionnaires are completed by patients to reflect their own perspective.[2,3]. Hip pain is a prevalent complaint, in which both the patient and the clinician could benefit from utilizing a PRO to monitor conditions and decide on a management approach.[4-5] The Harris Hip Score is a widely used tool which combines the clinicians input with the patient-reported symptoms to generate a better clinical picture of the hip pathology at hand and evaluate treatment options.[6] The questionnaire itself however, is in English. Healthcare services in Arabic speaking countries would not be able to use it, hence, the need for a cross-cultural adaptation of the score. The authors of this study aim to prove the validity and reliability of the Arabic version of this score.

Methods

Translation

The translation process was done as per recommendations of Guillemin's guidelines for validation and cross- cultural adaptation [9] after permission obtained from the original HHS copyright holder. Two Bilingual orthopaedic surgeons were responsible for the conceptual and literary translation of the original

HHS. Two other versions were produced by independent translation companies with a background in scientific English. All the versions produced were similar. Modifications to incorporate from all the version were made and implemented in the final version. A professional Arabic grammar checker reviewed it. The back-translation came close to the original score. A pilot test was then conducted on 10 random patients from the arthroplasty clinic after the Arabic version was approved by the translation committee. Both the physicians interviewed the patients after completing the questionnaire to address any issues or need for assistance.

Participants

110 patients completed the Harris Hip Score questionnaire, and agreed to have their data analyzed for research purposes. Average age of the participants was 44.3 years, with standard deviation of 15.4 years, implying that the majority of the sample was between 30 and 60 years of age. The youngest participant was 16, and the oldest was 76 years of age.

Psychometric Properties and Data Analysis

For all of the analyses IBM SPSS Statistics 21 was used.

In order to estimate reliability of the questionnaire Cronbach's alpha was calculated, and since every patient completed the survey on three different occasions, Cronbach's alpha was calculated for each of the three test situations. Also, ICC (interclass correlation coefficient) was used to assess test-retest reliability.

Content validity was tested by examing the shape of data distribution, as well as floor and ceiling effects. Floor effect is the percentage of patients who scored the lowest possible score (score of 0), and ceiling effect is the percentage of those with the highest score (score of 100). If more than 30% of the respondents had a floor or ceiling effect, the effects would be considered to be relevant.

To test convergent validity of the HHS, Spearman's correlation coefficient between HHS and WOMAC was calculated. Since WOMAC has already been validated in Arabic speaking countries, higher correlation coefficient would prove convergent validity of the HHS. Nonetheless, it is important to note that higher score on WOMAC indicates greater disability, while patients with greater disability would score low on HHS. This means that in order to have HHS validated, we are to expect negative correlation between score on WOMAC and HHS.

Questionnaires

Harris Hip Score

The HHS usually contains 12 questions covering four domains: pain, function, deformity, and range of motion. Questions are answered using a Likert scale, with the final score having a maximum of 100 points (best possible outcome), and minimum of 0 points (extreme symptoms). Those 100 points are divided into subdomains - pain receives 44 points, function 47 points, range of motion 5 points, and deformity 4 points; function is subdivided into activities of daily living (14 points) and gait (33 points). A total HHS of <70 points is considered to be a poor result, 70 to 80 is fair, 80 to 90 is good, and 90 to 100 is excellent (Nilsdotter and Bremander, 2011). For the purposes of this study a modified HHS - which is subtracted from the deformity and range of motion subdomains - is used. Hence, the possible range for this instrument is not from 0 to 100, but from 0 to 91. This means that ceiling effect would be documented for those patients who had scored 91 points.

All 110 patients have completed HHS in at least two different occasions (T1 and T2), and 109 of them completed a the third time (T3). There were two and a half weeks between each of these three occasions.

Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[8]

24 Likert-type items make this WOMAC, and using it, every patient gets three scores, from three different subscales. First subscale – pain – has 5 questions (score range 0-20), 2 questions address stiffness (score range 0-8), and physical function has 17 questions (range 0-68). A 0 score on each of the subscales means that patient essentially has not felt any discomfort in his/her hip (if any); on the other hand, a higher score suggests greater disability.

The survey was taken in two different occasions, and 2 weeks had passed between the two testing situations.

Results

WOMAC questionnaire

WOMAC has been validated in Arabic speaking countries, and has since been used in practice. Nevertheless, additional analyses have been conducted in order to explore psychometric characteristics of a WOMAC questionnaire that had been used in this study.

Test reliability for the first testing situation - 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. For the second testing, reliability was 0.99, 0.97, and 0.99 (pain, stiffness, and physical function, respectively). This only proves that WOMAC is an instrument with good reliability.

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 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. Being that these percentages are far less than 30% (which is considered relevant) – this is an argument in favour of content validity of WOMAC.

Harris Hip Score

To test the reliability of the instrument, Cronbach's alpha was calculated. For each of the three testing occasions the reliability was very good or excellent – $\alpha_1 = 0.92$, $\alpha_2 = 0.91$, and $\alpha_3 = 0.90$. Intra-class correlation coefficient was good with the score of 0.76 (95% CI 0.44-0.88).

Floor effect was recorded for 1% of the patients, and 2% showed a ceiling effect in the first week of testing. Two and a half weeks later, 1% of respondents again showed ceiling effect, and there was no floor effect recorded. On the third testing, 1% recorded floor effect, and one more time ceiling effect hasn't been documented. Shapiro-Wilk test was used to check if the data significantly deviates from the normal distribution, and it showed that it did, in all three testing occasions.

Table 1. Descriptive statistics of Harris Hip Score questionnaire

		N[1]	Min[2]	Max[3]	Mean	SD[4]	Sk[5]	Ku[6]	Floor effect	Ceiling effect
HHS	**Week 1**	110	0	91	66.0	17.613	-1.232	1.494	1%	2%
	Week 2	110	0	87	61.1	17.841	-1.024	.692	1%	0%
	Week 3	108	0	85	52.6	18.563	-.565	-.015	1%	0%

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

A 2 weeks' test-retest reliability of HHS was applied to the present manuscript. Of the 110 patients that fulfilled the questionnaire, 108 responded to the second assessment after the initial evaluation.

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

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)
Physical Function	53.31	16.39	62.91	18.60			9.60	0.623 (0.262 - 0.793)	0.768 (0.416 - 0.884)
HHS	72.55	19.35	67.12	19.61	57.81	20.40	-14.74	0.755 (0.442 - 0.876)	0.902 (0.704 - 0.955)

** Minus sign in HHS means that the condition of patient has been worsened over time (lower score = Deterioration) / Plus sign in WOMAC means that the condition of patient has been worsened over time (higher score = Deterioration)*

Test-retest reliability was also performed using Intra-class Correlation (ICC). The results (Table 2) indicated that HHS has an acceptable intra-class correlation with 0.755 (95% CI 0.442, 0.876). Considering the value of 0.902 (95% CI 0.704 – 0.955) for Cronbach's alpha, the internal consistency of the three assessments were proven to be very high.

In order to be able to compare the results of WOMAC questionnaire with those from HHS, it was important to standardize the scores of WOMAC to the range of 0-100. In addition, HHS score which were in the range of 0-91, 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 HHS decreased which is related to more pain and symptoms. At the same time the WOMAC mean score is showing an upward trend, which is also related with more pain and in general worsened conditions of the patient. This illustrates a visual agreement between the two questionnaires.

As can be seen in the table below, there are medium to large negative correlations between Harris Hip 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 low scores on HHS. This means that those who experience greater hip pain have higher scores on WOMAC, and lower HHS.

Table 2. Convergent validity of the Harris Hip Score (Spearman's rank correlation coefficient)

	WOMAC		
	Pain	Stiffness	Physical function
Week 1			
Harris Hip Score	-.56**	-.61**	-.62**
Week 2			
Harris Hip Score	-.41**	-.42**	-.48**

Note: ** Correlation is significant at the 0.01 level (2-tailed).

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%).

Table 3: Responsiveness and agreement between the two questionnaires

QUESTIONNAIRES		HRRIS HIP SCORE (HHS)			TOTAL
		Stable	Improvement	Deterioration	
WOMAC	Stable	**3.7%**	2.8%	30.8%	**37.4%**
	Improvement	0.0%	2.8%	10.3%	**13.1%**
	Deterioration	3.7%	0.9%	44.9%	**49.5%**
TOTAL		**7.5%**	**6.5%**	**86.0%**	**100.0%**

On the other hand, only 8 patients (7.3%) reported to remain stable by responding to HHS questionnaire. The majority of them (86.4%) believed their condition to be deteriorated, and only 6.4% of them reported relevant improvement after 2 weeks. In addition, it is relevant to note that 12 patients (11.2%) showed contradictory results (1 patient improved according to HHS and worsened according to WOMAC, while 11 patients showed the opposite). 33 patients (30%) believed that their condition was worsened according to HHS, while according to WOMAC their conditions were not changed (Table 3).

Table 4: Effect Sizes and SRMs for the WOMAC subscales and HOOS subscales. Bars represent the 95% confidence intervals.

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
HHS		0.729	0.537	0.891	0.456	0.441	0.467

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

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.20 According to responsiveness test, WOMAC subscales show similar responsiveness (SRM = 0.41) between first and second measurement. In comparison to WOMAC, HHS showed better responsiveness with SRM

= 0.46. 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 HHS. As shown in the graphs, overall mean difference between WOMAC and HOOS shows that there could be a systemic bias between two questionnaires (M = -7.49, 95% CI -13.59, -1.41, p = 0.016). In order to test this result, linear regression was performed with mean difference between WOMAC and HOOS as a dependent variable and mean value of WOMAC and HOOS as independent variable. The result of linear regression also indicates statistically significant difference between the two measurement methods (β = -0.94, 95% CI -1.801 – -0.081, t = -2.168, p = 0.03).

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 bias between the two methods (M = -18.9, 95% CI -25.13, -12.65, p < 0.001), the performed linear regression also confirms this bias (β = -0.95, 95% CI -1.81 – -0.104, t = -2.235, p = 0.028). It means that HHS 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 = -0.58, 95% CI -1.38 – 0.23, t = -1.429 p = 0.156).

Discussion

The primary objective of this study was to create a reliable and valid Arabic version of HOOS by translation and adaptation. For this purpose, the Arabic version of HHS is compared to the efficacy and results of WOMAC questionnaire. Preliminary validity and reliability tests revealed that there is moderate reverse correlation between WOMAC subscales and HHS, which indicated that they are related in the right direction, since their scores are in the opposite directions (0 for WOMAC = no pain / 0 for HHS = extreme pain).

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[18.] 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 HHS. The Bland-Altman plots indicated that there is systemic bias between WOMAC and HHS. And the linear regression illustrated that with increasing mean score, Arabic version of HHS tends to underestimate the results of WOMAC. According to Mcgrory et. al.[19], Differences in scores between hips were highly correlated for HSS and WOMAC total score, HHS pain and WOMAC pain subscores, and HHS function and WOMAC physical function

subscores. However, they found out that WOMAC stiffness and HHS range of motion were not significantly correlated. Overall they concluded, that patients with bilateral hip arthroplasty can apply the WOMAC osteoarthritis index questions to individual hips at the same time as effectively as the joint-specific HHS questions. The illustrated forest plots, and effect sizes, showed that HHS scores 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, especially about the improvement or deterioration, however, it is important to be cautious using HHS when the change magnitude of patient's condition is investigated, since there is a potential probability that the level of improvement of the patient's condition will be overestimated by HHS.

The major outcome of this study is that the HSS Arabic version demonstrated high levels of validity and reliability of evaluated patient-reported outcomes of Arabic patients with a range of hip pathologies. The patients did not encounter any difficulty in completing the questionnaire. An evaluation of the internal consistency showed that Cronbach's α coefficient for the HSS Arabic version was within the recommended range of values [10], the implication being that the questionnaire items were nonredundant as well as homogenous. The Arabic version of the HSS appears to have an excellent test-retest reliability (ICC, 0.755), compared to data reported in previous literature [11]. Hinman et al reported lower test-retest reliability with a 0.76 ICC value which corresponds with ours [12]. Interval of time between repeat measurements is a vital issue to be considered when determining the reliability of test-retest. According to the literature, estimation of HSS test-retest reliability ranges from 7-14 days, and three weeks to a month [11,12]. If patients are given short-retest intervals, then there is the risk of them getting over-familiar with the questions, while answers given will depend on their potential to recall the answers in the first assessment. Although this possibility is decreased by longer intervals, one may observe a spontaneous improvement of acute complaints. Generally, there should be a very short period between repeat administrations of outcome measures reported by the patient, when the condition being measured is expected to undergo a rapid change. The test was repeated seven day after the initial assessment. Hinman et al did a ~7.5 day interval retest for the hip patients (7-14 days), which corresponds with our study [12].

Celik et al.[21] sought to translate and culturally adapt the HHS into Turkish, and thereby determine the reliability and validity of the translated version. Celik et al, translated the HHS into Turkish in accordance with Beaton-recommended stages. 80 patients were tested by the HHS. The Turkish version of the HHS showed sufficient internal consistency (Cronbach's alpha, 0.70) and test-retest reliability (ICC = 0.91) compared to the Arabic version which had a test-retest reliability of 0.755 [11]. The Turkish study' correlation coefficients between the WOMAC & the OHS, and the HHS were 0.89 and 0.64 respectively [21]. The highest correlations between the HHS and SF-36 were with the physical function scale (r = 0.72), and the lowest correlations were with the mental function scale (r = 0.10). Celik et al. observed no floor or ceiling effects.

The literature has reported several validity tests. Studies conducted recently have sought to investigate the validity of the HHS by determining the link that it has with other outcome measures reported by

patients, such as the Total Functional Score [13], the WOMAC [11, 14], and the Nonarthritic Hip Score [15]. Our study provided evidence for construct validity by establishing the link between the Arabic versions to the WOMAC. The Arabic version of the HHS and the WOMAC had a very good construct validity (r = 0.67), which corresponded with that in previously documented data [12, 16].

Evidence for discriminate validity and convergent validity was provided. We determine what links existed between the eight scale scores and the HHS and 2 summary scores of the SF-36. Of course, the HHS had a strong relationship with concurrent measures of physical function compared to concurrent measures of mental function. We found the lowest correlation value between the HHS and mental domains of the SF-36 (r = 0.014). This demonstrates that the SF-36 measures additional aspects of physical health and provides more comprehensive, but less specific, information about a patient's overall health than do condition-specific questionnaires.

Conclusion

The primary purpose of this study was to create a reliable and valid Arabic version of HHS by translation and adaptation. Its reliability - calculated both through Cronbach's alpha and ICC - was good or moderate. Although the distributions for all subscales deviate from a normal one, no significant ceiling or floor effects were observed.

The Arabic version of HHS is short and easily administered and interpreted with minimal investment of time required for both the researcher and clinician. It is our belief that the Arabic version of the HHS is sufficient to evaluate the state of a Hip disease. Its levels of reliability and validity are acceptable and we believe that it will facilitate assessment of functional limitations and symptoms experienced by Arab-speaking individuals with a variety of hip disorders. There is need for further studies to assess responsiveness and to determine the minimum clinically relevant differences in the Arabic version of the HHS for common Hip pathologies.

Declarations

Ethic approval and Consent for publication

ethical approval was obtained

consent of publication was obtained in written from all participant

name of ethical committee: Ministry of health in Kuwait, Researchh and publication office

committee reference number: 0096524622226 , 00965 24622230

Consent to publish

Consent of participation and publish was obtain with written Format from all participants

Availability of data and material:

The data that support the findings of this study are available from [ministry of health Al-raze hospital, Kuwait] but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of [ministry of health Al-razi hospital ,Kuwait].

Competing interests:

Not applicable

The authors declare that they have no competing interests

Funding:

Not applicable

no funding was supplied in this case report

Authors' Contributions

The data collection and the writing was done by all the four authors equally.

Acknowledgment:

Not applicable

References

1. Sørensen, N., Hammeken, L., Thomsen, J. and Ehlers, L. Implementing patient-reported outcomes in clinical decision-making within knee and hip osteoarthritis: an explorative review. BMC Musculoskeletal Disorders. 2019; 20(1).
2. Siljander, M. P., McQuivey, K. S., Fahs, A. M., Galasso, L. A., Serdahely, K. J., & Karadsheh, M. S. (2018). Current Trends in Patient-Reported Outcome Measures in Total Joint Arthroplasty: A Study of 4 Major Orthopaedic Journals. The Journal of Arthroplasty.

3. Noble PC, Dwyer M, Brekke A. Commonalities, differences, and challenges with patient-derived outcome measurement tools: function/activity scales. Clin Orthop Relat Res 2013; 471:3457e65.

4. Pynsent PB. Choosing an outcome measure. J Bone Joint Surg Br. 2001; 83(6):792-794.

5. Puig-Junoy J, Ruiz Zamora A. Socio-economic costs of osteoarthritis: a systematic review of cost-of-illness studies. Semin Arthritis Rheum. 2015;44:531–41

6. Harris W.H. 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(4):737–755.

7. Klassbo, E. Larsson, E. MannevikHip disability and osteoarthritis outcome score. An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. Scand J Rheumatol, 32 (2003), pp. 46-51

8. Nilsdotter, A., and Bremander, A. (2011) Measures of hip function and symptoms: Harris Hip Score (HHS), Hip Disability and Osteoarthritis Outcome Score (HOOS), Oxford Hip Score (OHS), Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH), and American Academy of Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire. *Arthritis Care & Research*, Vol.63, pp 200-207.

9. 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(12): 1417-1432.

10. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007; 60(1):34-42.

11. Söderman P, Malchau H, Herberts P. Outcome of total hip replacement: a comparison of different measurement methods. Clin Orthop Relat Res. 2001; (390):163-172.

12. Hinman RS, Dobson F, Takla A, O'Donnell J, Bennell KL. Which is the most useful patient-reported outcome in femoroacetabular impingement? Test-retest reliability of six questionnaires. Br J Sports Med. 2014; 48(6):458-463.

13. Wamper KE, Sierevelt IN, Poolman RW, Bhandari M, Haverkamp D. The Harris hip score: Do ceiling effects limit its usefulness in orthopedics? Acta Orthop. 2010; 81(6):703-707.

14. Naal FD, Sieverding M, Impellizzeri FM, von Knoch F, Mannion AF, Leunig M. Reliability and validity of the crossculturally adapted German Oxford hip score. Clin Orthop Relat Res. 2009;467(4):952-957

15. Christensen CP, Althausen PL, Mittleman MA, Lee JA, McCarthy JC. The nonarthritic hip score: reliable and validated. Clin Orthop Relat Res. 2003; (406):75-83.

16. Shields RK, Enloe LJ, Evans RE, Smith KB, Steckel SD. Reliability, validity, and responsiveness of functional tests in patients with total joint replacement. Phys Ther. 1995; 75(3):169-176, discussion 176-179.

17. Potter BK, Freedman BA, Andersen RC, Bojescul JA, Kuklo TR, Murphy KP. Correlation of Short Form-36 and disability status with outcomes of arthroscopic acetabular labral debridement. Am J Sports Med. 2005; 33(6):864-870.

18. Altman, D.G. and Bland, J.M. (1983). Measurement in medicine: the analysis of method comparison studies. Statistician; 32: 307–317

19. McGrory BJ, Harris WH. (1996). Can the western Ontario and McMaster Universities (WOMAC) osteoarthritis index be used to evaluate different hip joints in the same patient? J Arthroplasty, 11(7): 841-4.

20. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Medical Care 1990;28(7):632-42

21. Celik D, Can C, Aslan Y, Ceylan HH, Bilsel K, Ozdincler A. Translation, Cross-Cultural Adaptation, and Validation of the Turkish Version of the Harris Hip Score. Hip Int 2014(; 5:) 24473-479.

Figures

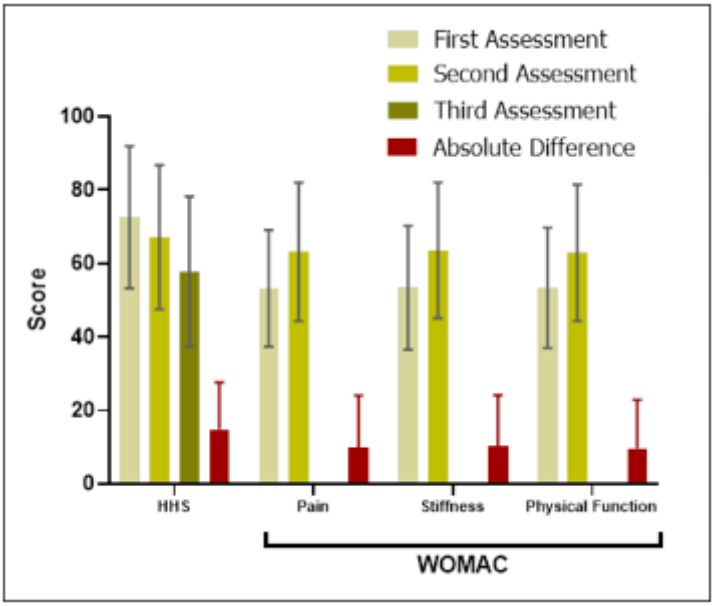


Figure 1

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

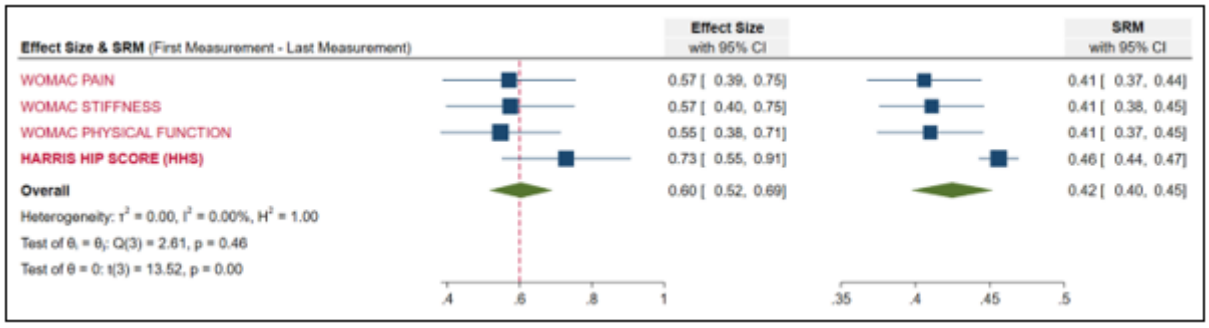


Figure 2

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

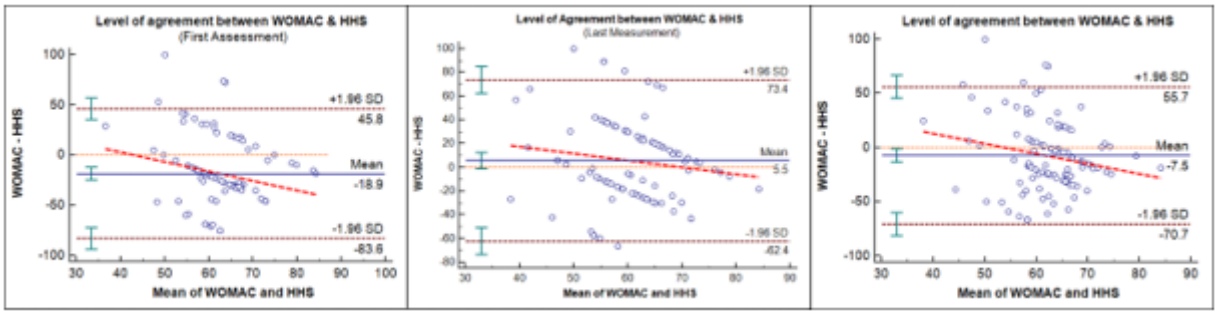


Figure 3

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