Journal of
Musculoskeletal Disorders and Treatment

ORIGINAL RESEARCH

Validity and Reliability of the Oxford Shoulder Instability Score Translated into Arabic

Sager Samir Hanna[1] (iD) *, Aliaa Fareed Khaja[1]** (iD) *, Ahmed Khaled Bouhamrah[1,2] and Ali Maqdes[1]* (iD)

[1]*Al-Razi Orthopedic Hospital, Kuwait*

[2]*Upper Extremity Fellowship, Queen's University, Canada*

*Corresponding author: *Aliaa Fareed Khaja, MD, Division of Orthopedic Trauma, Department of Orthopedic Surgery, Al-Razi Orthopedic Hospital, Sabah Medical Area, Kuwait City, Kuwait, Tel: +965-66135777*

Abstract

Background: The Oxford Shoulder Instability Score, abbreviated OSIS, is a brief, outcome measure self-reported by the patient suffering shoulder instability.

Objectives: Our objective was to translate OSIS into Arabic and validate its psychometric properties via test of the reliability, internal consistency, floor and ceiling effects, and validity.

Material & Methods: Fifty-five patients were involved in this survey at the baseline and follow-up (14 days after the baseline). We performed the internal consistency test using Cronbach's α. We calculated Standard Response Mean (SRM) and Pearson's Correlation to estimate the construct validity and responsiveness of the Arabic OSIS in comparison to Disability of the Arm and Shoulder and Hand (DASH) Score.

Results: The Arabic OSIS had a baseline Cronbach's α of 0.815 and a follow-up value of 0.860. In addition, Intra-class correlation (ICC) of 0.897; (0.813-0.942) indicated high reliability. Arabic versions of OSIS had a strong correlation with DASH score (r = 0.77, p = 0.003) which suggested a good construct validity. Also, moderately correlated changes of baseline to follow-up in OSIS indicated moderate responsiveness. We did not observe any relevant floor and ceiling effect among the responses.

Conclusion: Overall, the Arabic version of OSIS proved to be a good and reliable diagnostic tool for patients with shoulder instability.

Keywords

Validation, Arabic, Oxford, Shoulder, Instability, Score

Introduction

Shoulder instability is a common occurrence in orthopedics. It is most prevalent in young and physically active patients [1-3].

Evaluation of shoulder instability therapies should be assessed with outcomes that can be objectively verified, such as re-dislocations and range of motion, as well as subjective functioning. There is a range of patient-reported outcome measures (PROM) available for this purpose. Some are designed with the goal of capturing the patients' perspective of health and disease impact [4]. Because clinicians and patients do not readily agree on post-therapeutic physiological outcomes, PROM's have become important in the assessment of the health status of the patient [5,6]. Emphasis may be placed on the patient's general health, body part or physical domain (like the shoulder), or a specific condition, like instability [6-8].

The Oxford Shoulder Instability Score is a questionnaire comprising 12 questions. The questions are comprehensive and aimed at assessing the shoulder instability. The OSIS is a very important outcome measure in many clinical researches [9-11] but has yet to be translated into Arabic.

Translation of internationally applied PROM's as well as their validation will result in culturally equivalent instruments while permitting direct comparisons of international and national study results [12-14]. The objective of this study is translation and validation of the OSIS

Hanna et al. J Musculoskelet Disord Treat 2020, 6:082

• Page 1 of 5 •

for the Arabic population and the evaluation of its measurement properties according to current guidelines in the literature [15].

Disabilities of the Arm Shoulder and Hand (DASH) Score

The DASH score comprises 30 items. All items are self-reported and designed to measure physical symptoms and functions in patients experiencing musculoskeletal disorders of the upper limbs [16]. The objective of the DASH score is to describe the disability experienced by this group of patients and to monitor any changes of function and symptoms over time after treatments [17].

The DASH score has proven to be a reliable tool for the investigation of joints in the upper extremities. Each item is scored from 0-4 with the total score being calculated by summing the score of all rated items (0-120). The DASH score was used because it has already been validated in the Arabic Language.

Material & Methods

Translation

We did the translation as per recommendations of Guillemin's guidelines for validation and reliability after permission obtained from the original OSIS copyright holder [13]. Two bilingual orthopaedic surgeons were responsible for the conceptual and literary translation of the original version. 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 versions 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 ten random patients from the Sports Shoulder clinic. This was done after the approval of the Arabic version by the translation committee. Both the physicians interviewed the patients after completing the questionnaire to address any issues or need for assistance.

Participants

Fifty-five patients participated in this study and completed the OSIS and DASH scores and agreed to have their data analysed for research purposes. The youngest participant was 21, and the oldest was 35 years of age. The patients are Arabic-speaking patients that presented to the specialized shoulder clinic, which is the only clinic available in the public sector. All these patients have had two or more dislocations before presenting to this clinic.

Psychometric Properties & Data Analysis

Internal consistency

The outcome measures of each construct were pre-

sented using descriptive analysis. Mean and standard deviation (SD) were calculated. Internal consistency was evaluated by calculating the Cronbach's α. Internal consistency determines to what extent different items within one questionnaire measures the same construct of interest. According to the literature, $\alpha > 0.70$ is regarded as acceptable, while it should not be higher than 0.95, in order to avoid redundancy [18].

Reliability

The reliability refers to the proportion of the total variance in the measurements that can be attributed to true differences between patients [7]. Reliability was estimated by calculating the ICC, which was calculated with a two-way, mixed-effects model for absolute agreement, and scores larger or equal to 0.70 were considered adequate [19].

Construct validity

Construct validity determines whether the questionnaire measures what it was designed to measure. In the case of shoulder instability, do questions actually measure the typical complaints following shoulder instability? In order to investigate the construct validity of the Arabic OSIS, its relationship to a comprehensive questionnaire like the DASH score had to be examined. For this purpose, Pearson's correlation coefficient between Arabic OSIS and DASH was calculated. Since the DASH score had already been validated in Arabic speaking countries, higher correlation coefficient would prove convergent validity of the Arabic OSIS. Furthermore, content validity was measured by examining the floor and ceiling effects. Floor effect is the percentage of patients who scored the lowest possible score (score of 0), while ceiling effect is the percentage of those with the highest score (score of 48). If more than 15% of the respondents had achieved the highest or lowest score, then floor or ceiling effects would be present and this would limit the validity of the content of the questionnaire [20].

In addition, the responsiveness, which indicates how well a questionnaire shows clinically important changes over time, was measured by software MedCalc. To determine responsiveness of the Arabic version of OSIS, Standardized Response Mean (SRM) was also calculated.

The calculations were performed using IBM SPSS v.26, MedCalc v.19.1 and Graphpad Prism v.8.

Results

Fifty-five patients participated in this study and completed the OSIS and DASH scores and agreed to have their data analysed for research purposes. Average age of the participants is 27.18 years, with standard deviation of 4.29 years, which means that the majority of the sample was between 22.89 and 31.47 years of age. The youngest participant was 21, and the oldest was 35
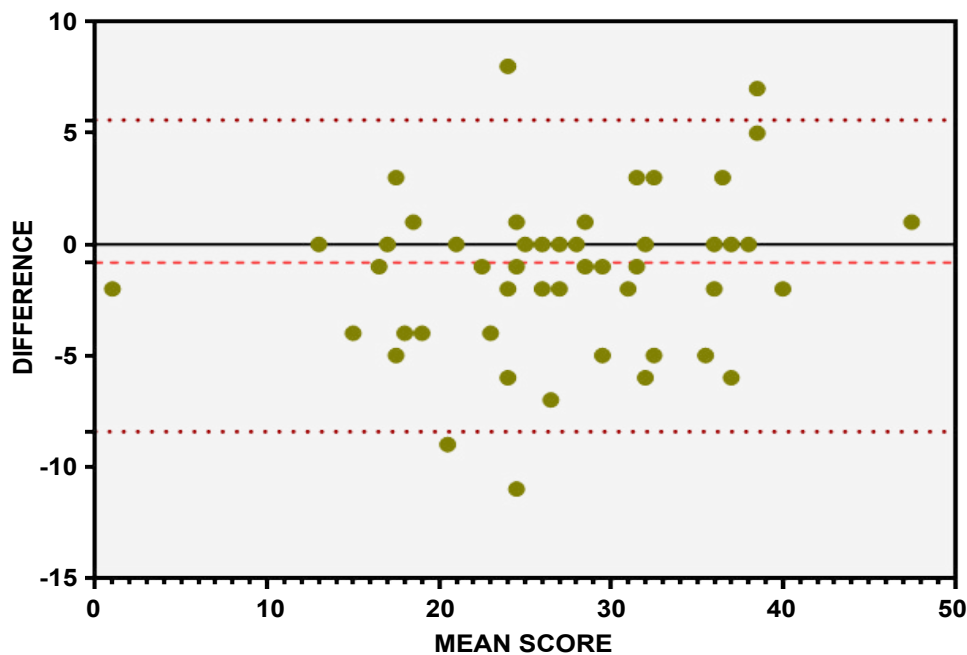
Figure 1: Bland-Altman plot to visualize the level of agreement between Test and Re-test of Arabic OSIS.

Table 1: Descriptive analysis of baseline and follow-up outcome measures.

	Baseline		Follow-up	
	Mean	SD	Mean	SD
Arabic DASH Score	66.36	20.21	-	-
Arabic OSIS	26.72	8.57	28.15	8.05

Table 2: Intra-class correlation and standard response mean (SRM) of Arabic OSIS.

Arabic OSIS	Mean	SD	ICC	95% CI	SRM
Baseline	26.72	8.57	0.897	0.813-0.942	0.69
Re-test	28.15	8.05			

years of age. Both ceiling and floor effect were recorded to be at 2%, which is not relevant. Table 1 illustrates the analysis of the scores completed by the participants at baseline and at follow-up. The mean time between the completion of the first and second questionnaires was 14 days.

Psychometric Analysis

Reliability & internal consistency

In order to estimate the reliability of the questionnaire, Internal consistency was calculated by using overall Cronbach's α which was equal to 0.815 at the baseline and 0.860 at follow-up, indicating a high degree of internal consistency in both time frames.

Table 2 presents the scores of the tests and re-tests and the ICC with a 95% confidence interval (ICC is 0.897; 0.813-0.942), which indicate excellent reliability. In Figure 1, Bland-Altman plot demonstrates the level of agreement between test and re-test of Arabic OSIS. The plot indicates that Arabic OSIS has a reliable replicability.

Responsiveness & construct validity

The Arabic versions of OSIS and DASH scores indicated a strong correlation between them (r = 0.774, p = 0.003). The strong correlation is an indication of strong construct validity. In addition, the SRM (Standard Response Mean) for Arabic OSIS was measured with SRM = 0.69, which was moderate.

Discussion

There is an increasing trend by institutions in utilizing PROM's, both for research and for clinical purposes, as it finds great application in supplementation of measures of clinical outcomes. With Cronbach's alpha valued at 0.92, and a 5.7 measurement error, OSIS has proven to be reliable and valid, thus proving its clinical importance to patients experiencing shoulder instability [21]. To our knowledge, our study is the first validation of OSIS in Arabic.

The internal consistency as indicated in the result is on the high side (Cronbach's α = 0.815 at the baseline, and 0.860 at the follow-up). This is just slightly below the value highlighted in the original article (Cronbach's α = 0.91 at the baseline [n = 92] and 0.92 at follow-up [n = 64]). In our study, no relevant floor and ceiling effect was observed among any of the responses.

Taking into consideration the context of the questions, it suffices to say that the OSIS effectively determines a number of constructs including pain, social-, physical-, and role functioning, as well as frequency of worries and dislocation.

We demonstrated the extent of agreement between the test retest of the Arabic OSIS with the Bland-Altman plot. From the plot, it is evident that Arabic OSIS is re-

Hanna et al. J Musculoskelet Disord Treat 2020, 6:082

• Page 3 of 5 •

liably replicable. Test-retest of the sample addressed the reliability within a mean time interval of 14 days resulting in an ICC of 0.897. This is lower than the 0.97 value described by Dawson, et al. after an interval of 24 hours in 34 patients [21]. That notwithstanding, an ICC of 0.897 is considered very good.

Dawson, et al. assessed the construct validity by calculating correlations with the Rowe & Constant scores. In addition, the Constant Score doesn't apply to shoulder instability [22,23]. Even though the DASH score does have a much broader range of functional questions none of these pertain specifically to instability. However, the DASH score was chosen to be used in this study as it was already validated in Arabic. We assessed construct validity by calculating Pearson's Correlation Coefficient between Arabic OSIS and the DASH. With a value of r = 0.774 (p = 0.003), we consider the construct validity to be good. This high correlation is more specific in addressing daily activities than the OSIS. This correlation may be compared to Dawson, et al., an indication that, alongside physical pain, the OSIS also measures aspects of role limitations and pain due to physical problems.

The OSIS was translated and validated into several languages. In 2015, a Dutch version of the OSIS was validated and evaluated for reliability [24]. In their study, 138 patients completed the Dutch version of the OSIS at baseline and a subgroup completed the follow-up retest at an average of 13 days. The internal consistency was measured using Cronbach's α, it was found to be 0.88 [24]. The reliability (ICC) was found to be excellent (0.87) [24]. Construct validity was evaluated by comparing OSIS with several outcome measures. Of note was the WOSI (with highest correlation with OSIS 0.82), and the DASH (0.79) [24]. They concluded that the Dutch OSIS showed good reliability and validity in patients with shoulder instability.

Olyaei, et al. produced a prospective cohort study of the Persian OSIS translation and validation [25]. Their study population was 150 patients. Internal consistency using Cronbach's α was 0.90 [25]. Test retest reliability (ICC) was shown to be excellent (0.94) [25]. They showed the Pearson correlation coefficient between the Persian OSIS and DASG which was 0.84 [25]. This indicated good convergent validity.

Mazzoni, et al. evaluated the reliability, validity, and reproducibility of an Italian version of OSIS (sample size 25 patients) [26]. Cronbach's alpha in their study was 0.897, while their ICC was 0.805 [26]. They concluded that the Italian OSIS is a reliable, valid and reproducible outcome measure for clinical evaluation.

The strength of our study was that the population size (n = 55) with no missing values. Conversely, a limitation of our study was the total number of questions assigned to the patients. Answering questions from different questionnaires requires a level of time and focus, and there is the possibility of the patients losing focus or digressing. Although electronic versions may have the advantage of high follow-up ratio and prevention of data misplacement, validation of digital formats is still important and should be carried out. Another limitation is that the paper compares the OSIS which is an instability score with the DASH score which is a score assessing general upper limb dysfunction. The study would have more strength if there was a comparison with another instability score in addition to the DASH score. Future studies will do well to specify the exact scoring system utilized.

Conclusion

In this study, we found that the OSIS Arabic version could be relied upon as an outcome measure in patients experiencing shoulder instability, with an ICC of 0.897, and a Cronbach's α of 0.815. Also, we considered the construct validity to be good. The OSIS comprises 12 questions, is user-friendly, and can be administered with ease. Also, OSIS is of utmost importance as a PROM in clinical practice, without floor or ceiling effects.

Declarations

Ethical approval and consent for publication

- Ethical approval was obtained
- Consent of publication was obtained in writing from all participants
- Name of Ethical Committee: Ministry of Health, Kuwait, Research and publication office
- Committee Reference Number: 2019/1060

Consent to publish

Consent of participation and publish was obtained 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-Razi Orthopedic 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 the Ministry of Health Al-Razi Orthopedic Hospital, Kuwait.

Competing interests

The authors declare that they have no competing interests.

Funding

No funding was supplied in this study.

Authors' contributions

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

Acknowledgment

References

1. Leroux T, Wasserstein D, Veillette C, Khoshbin A, Henry P, et al. (2014) Epidemiology of primary anterior shoulder dislocation requiring closed reduction in Ontario, Canada. Am J Sports Med 42: 442-450.

2. Liavaag S, Svenningsen S, Reikeras O, Enger M, Fjalestad T, et al. (2011) The epidemiology of shoulder dislocations in Oslo. Scand J Med Sci Sports 21: e334-e340.

3. Zacchilli MA, Owens BD (2010) Epidemiology of shoulder dislocations presenting to emergency departments in the United States. J Bone Joint Surg Am 92: 542-549.

4. Haywood KL (2006) Patient-reported outcome: Measuring what matters or just another paper exercise? Musculoskeletal Care 4: 63-66.

5. Janse AJ, Gemke RJ, Uiterwaal CS, van dT I, Kimpen JL, et al. (2004) Quality of life: Patients and doctors don't always agree: A meta-analysis. J Clin Epidemiol 57: 653-661.

6. Wright RW, Baumgarten KM (2010) Shoulder outcomes measures. J Am Acad Orthop Surg 18: 436-444.

7. Irrgang JJ, Lubowitz JH (2008) Measuring arthroscopic outcome. Arthroscopy 24: 718-722.

8. Poolman RW, Swiontkowski MF, Fairbank JC, Schemitsch EH, Sprague S, et al. (2009) Outcome instruments: Rationale for their use. J Bone Joint Surg Am 91: 41-49.

9. Steffen V, Hertel R (2013) Rim reconstruction with autogenous iliac crest for anterior glenoid deficiency: Forty-three instability cases followed for 5-19 years. J Shoulder Elbow Surg 22: 550-559.

10. Tan CK, Guisasola I, Machani B, Kemp G, Sinopidis C, et al. (2006) Arthroscopic stabilization of the shoulder: A prospective randomized study of absorbable versus nonabsorbable suture anchors. Arthroscopy 22: 716-720.

11. van der Linde JA, van Kampen DA, Terwee CB, Dijksman LM, Kleinjan G, et al. (2011) Long-term results after arthroscopic shoulder stabilization using suture anchors: An 8- to 10-year follow-up. Am J Sports Med 39: 2396-2403.

12. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 25: 3186-3191.

13. Guillemin F, Bombardier C, Beaton D (1993) Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. J Clin Epidemiol 46: 1417-1432.

14. Wild D, Grove A, Martin M, Eremenco S, McElroy S, et al. (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 Health 8: 94-104.

15. Sousa VD, Rojjanasrirat W (2011) Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. J Eval Clin Pract 17: 268-274.

16. Hudak PL, Amadio PC, Bombardier C (1996) Development of an upper extremity outcome measure: The DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). Am J Ind Med 29: 602-608.

17. Kennedy CA, Beaton DE, Smith P, Van Eerd D, Tang K, et al. (2013) Measurement properties of the QuickDASH (disabilities of the arm, shoulder and hand) outcome measure and cross-cultural adaptations of the QuickDASH: A systematic review. Qual Life Res 22: 2509-2547.

18. Fayers PM, Machin D (2013) Quality of life: The assessment, analysis and interpretation of patient-reported outcomes. John Wiley & Sons, Manhattan.

19. Snyder CF, Aaronson NK, Choucair AK, Elliott TE, Greenhalgh J, et al. (2012) Implementing patient-reported outcomes assessment in clinical practice: A review of the options and considerations. Qual Life Res 21: 1305-1314.

20. McHorney CA, Tarlov AR (1995) Individual-patient monitoring in clinical practice: Are available health status surveys adequate? Qual Life Res 4: 293-307.

21. Dawson J, Fitzpatrick R, Carr A (1999) The assessment of shoulder instability. The development and validation of a questionnaire. J Bone Joint Surg Br 81: 420-426.

22. Jensen KU, Bongaerts G, Bruhn R, Schneider S (2009) Not all Rowe scores are the same! Which Rowe score do you use? J Shoulder Elbow Surg 18: 511-514.

23. Lillkrona U (2008) How should we use the Constant Score? -A commentary. J Shoulder Elbow Surg 17: 362-363.

24. van der Linde JA, van Kampen DA, van Beers LW, van Deurzen DF, Terwee CB, et al. (2015) The Oxford Shoulder Instability Score; validation in Dutch and first-time assessment of its smallest detectable change. J Orthop Surg Res 10: 146.

25. Olyaei G, Mousavi S, Montazeri A, Malmir K (2016) Translation and validation study of the Persian version of the Oxford shoulder instability score. JMR 10: 24-28.

26. Mazzoni B, Cucchi D, Giovannelli T, Paci M, Arrigoni P, et al. (2019) Translation, cross-cultural adaptation, and validation of the Italian version of the Oxford Shoulder Instability Score. Int Orthop 43: 2125-2129.

CLINMED
INTERNATIONAL LIBRARY