Original Article
IASR Multidisciplinary Journal
Website: http://iasrpublication.com/imj
ISSN (*online*): 2583-0988

Reliability and Validity of a pictorial version of the Oxford Shoulder Score

Husna Moola[1*]; Neil Kruger[1]; Stephen Roche[1]; Juan Klopper[1, 3]; Roan De Wet, Michael FG Held[1, 2]
[1]Department of Orthopaedic Surgery, Groote Schuur Hospital, Orthopaedic Research Unit, University of Cape Town, South Africa
[2]Knee Unit, Groote Schuur Hospital and Christiaan Barnard Memorial Hospital, University of Cape Town, South Africa
[3]Department of Surgery, Groote Schuur Hospital, University of Cape Town, South Africa

A B S T R A C T

Background: Patient reported outcome measures assist in assessing surgical outcomes. Most African countries have a high illiteracy rate. The aim of this study was to develop and validate a pictorial version of the Oxford Shoulder Score (OSS) against the English OSS and Simple Shoulder Test (SST). **Methods**: A pictorial version of the OSS was developed. There were 12 questions, each modelled on the English OSS. Wong Baker faces were used as a scoring scale. The pictorial score was validated on patients who had shoulder pathology excluding instability at a tertiary center in South Africa between May 2016 and November 2017. Participants were first asked to complete the Pictorial OSS and then the English OSS and SST. **Results**: 87 participants completed the scores. The pictorial score was internally consistent with a Cronbach's alpha of 0.95(CI 0.93-0.97). The Spearman Rank correlation coefficient was 0.8. The mean time required to complete the score, used as a surrogate marker of understanding, was 134s (STD 67s) which is less than the time taken to complete the English OSS or SST. **Conclusion**: The pictorial version of the OSS was reliable, reproducible, valid and acceptable, and can therefore be used in our patient population.

Keywords*: Pictorial, PROM (patient reported outcome measure), shoulder function, Oxford Shoulder Score (OSS).*

Corresponding Author
Husna Moola
Department of Orthopaedic Surgery, Groote Schuur Hospital, Orthopaedic Research Unit, University of Cape Town, South Africa

INTRODUCTION

Safety and outcomes of surgery in low-middle income settings such as Africa need to be assessed. Current guidelines are predominantly derived in European and North American populations limiting their applicability in our setting [1]. Patient reported outcome measures (PROMs) are important in assessing surgical outcomes. Appropriate standardized assessment of our patients also improves our ability to collaborate in international research efforts. Recently, the Oxford Shoulder Score has been translated into Afrikaans, a language uniquely spoken in Southern Africa [2]. However, many African countries have a high illiteracy rate with only 60% of Africans over the age of 15 years being able to read and write [3]. Furthermore, approximately 2000 languages are spoken in Africa [4]. South Africa alone has 11 official languages and 78% of school children are illiterate at a time where literacy would be expected [5]. Similar challenges are present in most low to middle income countries which makes the administration of English PROMs ineffective and translation an exhaustive and possibly futile task.

Pictorial information could overcome the language and literacy barriers patients face in completing PROMs and would therefore be globally applicable. The current evidence for picture-based scores is sparse. One study evaluated patients' pictorial assessment of menstrual blood loss compared to true blood loss [6]. Another preliminary study aimed to convert existing scores i.e. the EQ-5D-3L and EQ-5D-5L into pictorial scores and, although not validated yet, found promising results [7]. To date, no pictorial orthopaedic outcome score has been published.

Our study developed a pictorial PROM that evaluates shoulder function in patients. The pictorial score is based on the Oxford Shoulder Score. This pictorial version of the OSS was then compared to previously validated scores; the English OSS [8, 9] and Simple Shoulder Test [10] in order to assess its validity, reproducibility and user-friendliness.

MATERIALS AND METHODS

This validation study was conducted at a tertiary care facility in South Africa. Participants were recruited from May 2016 to November 2017. Patients with shoulder pathology and fluency in reading and understanding English were included. Participants were excluded if they were younger than 18 years of age, had shoulder instability or impaired vision precluding from reading. Patients with shoulder instability were excluded as the Oxford Shoulder Score was not initially validated in patients with recurrent dislocation or subluxation [9]. Participants who had surgical interventions

were excluded from the follow-up group. Fluency in reading and understanding English was self-declared by the participants and was not formally assessed.

Design of the pictorial score

The Pictorial Version of the Oxford Shoulder Score (OSS) consisted of twelve questions, each modelled on a specific question of the OSS. The pictures were commissioned by the research team and reviewed by a panel consisting of the research team and physiotherapists. Each question was transferred into three pictures which collectively depict an activity (Appendix 1). The Wong Baker Faces Scale was used as a scoring technique [11]. This scoring scale consisting of 6 faces displaying progressive increase in difficulty or pain (Appendix 2). This scale was chosen as it is consistently found to be preferred over other numeric, visual analogue and photographic scales [11].

After written consent was taken, the study was explained verbally to the participant. The self-administered questionnaire included the pictorial version of the OSS, as well as the English versions of the written OSS and Simple Shoulder Test (SST). The pictorial OSS was completed first. This was to prevent bias in understanding the tasks depicted by the pictures that could have occurred if patients completed the English OSS first. To assess reproducibility, a subset of the patients were followed up between 3 days and 3 months after their initial score ensuring no events such as trauma or surgery changed their shoulder function. The exact time is not recorded. They were then asked to repeat the pictorial version of the OSS. The retest was done at the next clinically indicated follow up and times between tests therefore varied to some degree.

Sample size

Our sample size was 87 participants. The retesting was assessed with a subset of 30 participants. Calculated with the OSS, a sample size of 19 achieves 80% power to detect a mean of paired differences of 4.5 with a known standard deviation of differences of 7.0 and with a significance level (alpha) of 0.05 using a two-sided paired z test. A mean of paired differences of 4.5 is clinically relevant for the Oxford Shoulder Score [8].

Statistical methods

The Mathematica statistical package was used (Wolfram Research, Inc., Mathematica, Version 11.3, Champaign, IL 2018.

Demographic data was analyzed with descriptive statistics. Continuous variables with normal distribution were compared using Students T test otherwise the Wilcoxon rank-sum test was used. Categorical variables were compared with the Chi-Squared test. A confidence interval of 95% was used and alpha was set at 0.05 to describe significant difference.

The pictorial version of the OSS was validated against the written version of the English OSS and the SST. Each of these scores has different scales and therefore a different maximum score. To assess external validity, they were compared using the Kruskal-Wallis coefficient, after normalizing the respective minimum and maximum values. The Kruskal Wallis significance value is 0.21. Internal validity was assessed using Cronbach alpha. The reproducibility was assessed by comparing the initial pictorial score with its retest. A Spearman Rank was calculated for each question as well as the average of all twelve questions.

Time taken to complete the score was used as a surrogate marker for understanding. The mean time needed to complete the score was calculated for each of the scores. For test-retest reliability, we defined a spearman correlation coefficient of 0,8 – 1 as very good/ strong, 0.6-0.79 as good/strong, 0.4-0.59 as moderate, 0.2-0.39 as weak and 0-0.19 as very weak [12].

Missing data

Missing data constituted less than 5% of data collected and was missing at random (not confined to particular variables). They were replaced by imputation of median values as proposed in the guidelines for validating the score [8].

Ethics

Approval for this study was given by the institution Human Research Ethics Council (HREC 340/2016 number). This study was conducted in accordance with the principles in the Declaration of Helsinki. Patients gave informed written consent prior to enrolment in the study.

RESULTS

87 patients were included in the study. All patients completed the three sections of the questionnaire i.e. the pictorial OSS, English OSS and the SST. All the questionnaires were self-completed. Most patients were middle-aged, with a mean age of 54 years (STD 14 years). 39 (44%) patients were male. Education level was divided into primary, secondary

or tertiary education. Most patients had secondary schooling or above. Table 1 demonstrates the demographics of the population.

Table I shows the demographics of the population.

Validity

Validity was assessed by comparing the mean score of the pictorial OSS to the written OSS and the SST (Table ii). As each score was assessed on a different scale, this was done after normalizing the data as seen in Figure i. The Kruskal-Wallis co-efficient is 0.12 (p 0.9)

Internal consistency and Reproducibility

Internal consistency was assessed using Cronbach alpha. The Cronbach alpha was 0.95 (CI 0.93-0.97). 30 patients filled in the questionnaire twice to assess reproducibility. Participants were followed up between 3 days and 3 months later. The overall Spearman rank was r=0.8 (p<0.5). The Spearman rank co-efficient for each question is shown in table iii.

Understanding

We used the time needed to complete each section of the questionnaire as a surrogate marker of understanding. The mean time required for completing the pictorial OSS was less than time needed for the English OSS and SST (Pictorial OSS 134s (STD 67) vs English OSS177s (STD 67)/ SST165 (STD 75))

DISCUSSION

The demographics and level of education was comparable to other validation studies (Table 4). The mean age of the population was 54 years (IQR 15) which is similar to other study populations [2, 13, 14 & 15]. There was a female preponderance in this study with 47 female and 40 male participants. This is similar to the Afrikaans, Turkish and German validation studies [2, 14 & 16]. The Dutch and Italian studies had a male preponderance [13, 15]. The pictorial version of OSS is reliable, reproducible, acceptable and valid. The pictorial OSS was externally valid with a Kruskal-Wallis Co-efficient of 0.12 (p 0.9). The pictorial score has a Cronbach alpha of 0.95 (CI 0.93-0.97) confirming internal validity. This is comparable to other validation studies as outlined in table iv.

The reproducibility of the pictorial OSS was assessed using Spearman Rank and was 0.8 (p<0.5). Other validation scores have higher Pearson Correlation Coefficients of 0.98 German, 0.97 Italian, 0.981 Dutch [13, 14 & 15]. Although a spearman correlation coefficient of 0.8 is still defined as a very good correlation, the lower retest reproducibility in this study could be due to a longer time interval between the two questionnaires. The exact time of retesting in was not recorded. The minimum of 72 hours between the first questionnaire and a retest is comparable to other studies; 24-72 hours for the German OSS, 48 to 96 hours for the Portuguese OSS, two days for the Romanian OSS [17, 18]. However, the maximum time interval, in the absence of trauma or surgery was 3 months. This could have resulted in small changes in shoulder function in the interim.

Time taken to complete the score was used as a surrogate marker of understanding and therefore acceptability of the pictorial OSS. The time taken to complete the pictorial OSS score was 134s. This is shorter than the time taken in other validation studies; 205s for the German and 480s for the Afrikaans [2, 14]. It is also shorter than the time taken by the same study population to complete the English OSS. This confirms that the pictures are easy to interpret in our population.

The mean score for the pictorial OSS is 46 (STD 17). This is higher than other validation studies on the same scale e.g. German (27.34 (STD 10.42)), Dutch (32.5 (STD 9.5)), Italian (36.05 (STD 13.95)). As a low to middle income country, patients may have inferior access to care and seek medical help later in disease processes when symptoms are more severe.

The limitations of this study are that it was carried out in a literate population, fluent in reading English. This was done to ensure that the English version of the OSS and the SST would be self-administered questionnaires. However, we believe that the characteristics of the study population are otherwise similar to the illiterate and non-English speaking population for which the pictorial score is intended. Another limitation is that this score was tested in an urban setting and cultural differences in rural areas might influence patients' interpretations of the pictures. We attempted to limit this shortcoming by including members in the review panel of the score who are familiar with rural life. The transferability to other countries might be limited as the score was only validated in our country, but the pictures are easily understandable and can therefore be used globally. The exact time of retesting was not recorded, although the time frame is known (3-90 days). However, the study was not designed and powered for a subgroup analysis within the retested participants to interrogate the effect of time on scores. Overall, the correlation was very good (0.8), which shows the time between tests

was acceptable. We did not collect information on shoulder pathology diagnoses or hand dominance. However, the inclusion criteria for this study was a population in which the OSS in previously validated i.e. shoulder pathology excluding instability. Also, the pictorial OSS score was only compared to the English OSS and the SST and not to further scores. But the focus of our study was to evaluate patient understanding of pictorial questions, therefore comparable participant scores in the English OSS and the pictorial OSS were used to assess the validity of the pictorial OSS.

CONCLUSION

The pictorial version of the OSS was reliable, reproducible, valid and acceptable. Using only images to assess shoulder function, it could be used to overcome the existing challenges of written scores with illiteracy, as well as language and cultural differences. Further research should increase the geographic use, as well as the design of pictorial outcome scores for other areas of disease.

Conflict of interest

The authors declare that there is no conflict of interest.

REFERENCES

1. Dare, A. J., Grimes, C. E., Gillies, R., Greenberg, S. L., Hagander, L., Meara, J. G., & Leather, A. J. (2014). Global surgery: defining an emerging global health field. *The Lancet*, *384*(9961), 2245-2247.
2. Kruger, N., Stander, L., Maqungo, S., Roche, S., & Held, M. (2018). The Oxford Shoulder Score: Cross-cultural adaptation and translational validation into Afrikaans. *SA Orthopaedic Journal*, *17*(1), 17-23.
3. UNESCO (2018). Literacy rate, adult total (% of people ages 15 and above) [Internet]. *The World Bank Data*. [cited 2019 Aug 6]. p. 1. Available from:https://data.worldbank.org/indicator/SE.ADT.LITR.ZS?end=2010&locations=ZG-1W-Z4-8S-Z7-ZJ&start=2010&view=bar
4. Ranking A. Top 10 most spoken languages in Africa [Internet]. 2016 [cited 2019 Aug 6]. p. 1–5. Available from: http://www.africaranking.com/most-spoken-languages-in-africa/
5. Index Mundi (2018). South Africa Literacy - Demographics [Internet]. *South Africa Literacy*. [cited 2018 Jun 16]. p. 1. Available from: https://www.indexmundi.com/south_africa/literacy.html
6. Reid, P. C., Coker, A., & Coltart, R. (2000). Assessment of menstrual blood loss using a pictorial chart: a validation study. *BJOG: An International Journal of Obstetrics & Gynaecology*, *107*(3), 320-322.
7. Whitehurst, D. G., Latimer, N. R., Kagan, A., Palmer, R., Simmons-Mackie, N., Victor, J. C., & Hoch, J. S. (2018). Developing accessible, pictorial versions of health-related quality-of-life instruments suitable for economic evaluation: a report of preliminary studies conducted in Canada and the United Kingdom. *Pharmaco Economics-open*, *2*(3), 225-231.
8. Dawson, J., Rogers, K., Fitzpatrick, R., & Carr, A. (2009). The Oxford shoulder score revisited. *Archives of orthopaedic and trauma surgery*, *129*(1), 119-123.
9. Dawson, J., Fitzpatrick, R., & Carr, A. (1996). Questionnaire on the perceptions of patients about shoulder surgery. *The Journal of bone and joint surgery. British volume*, *78*(4), 593-600.
10. Lippitt S.B., Harryman D.T., and Matsen F.A. (1993): A practical tool for evaluating function: the Simple Shoulder Test. In Matsen F.A., Fu F.H., and Hawkins R.J. (eds): *The shoulder: a balance of mobility and stability*. Rosemont (IL): *American Academy of Orthopaedic Surgeons*.
11. Wong, D. L., & Baker, C. M. (2001). Smiling face as anchor for pain intensity scales. *Pain*, *89*(2), 295-297.
12. Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal*, *24*(3), 69-71.
13. Berendes, T., Pilot, P., Willems, J., Verburg, H., & te Slaa, R. (2010). Validation of the Dutch version of the Oxford Shoulder Score. *Journal of shoulder and elbow surgery*, *19*(6), 829-836.
14. Huber, W., Hofstaetter, J. G., Hanslik-Schnabel, B., Posch, M., & Wurnig, C. (2004). The German version of the Oxford Shoulder Score—cross-cultural adaptation and validation. *Archives of orthopaedic and trauma surgery*, *124*(8), 531-536.
15. Murena, L., Vulcano, E., D'Angelo, F., Monti, M., & Cherubino, P. (2010). Italian cross-cultural adaptation and validation of the Oxford Shoulder Score. *Journal of shoulder and elbow surgery*, *19*(3), 335-341.
16. Tuğay, U., Tuğay, N., Gelecek, N., & Özkan, M. (2011). Oxford Shoulder Score: cross-cultural adaptation and validation of the Turkish version. *Archives of orthopaedic and trauma surgery*, *131*(5), 687-694.
17. Gonçalves, R. S., Caldeira, C. Q., Rodrigues, M. V., Felícia, S. C., Cavalheiro, L. M., & Ferreira, P. L. (2018). Cross-cultural adaptation and validation of the Portuguese version of the Oxford Shoulder Score (OSS). *Acta Reumatol Port*, *43*(2), 102-8.
18. Haragus, H., Prejbeanu, R., Patrascu, J., Faur, C., Roman, M., Melinte, R., ... & Marra, G. (2018). Cross-cultural adaptation and validation of the Romanian Oxford Shoulder Score. *Medicine*, *97*(23).