

RESEARCH

Open Access



# Correlation and responsiveness of global health, upper extremity-specific, and shoulder-specific functional outcome measures following reverse total shoulder arthroplasty for proximal humerus fracture

James Barger<sup>1,2\*</sup>, Dafang Zhang<sup>1,2</sup>, Derek S. Stenquist<sup>1,2</sup>, Peter Ostergaard<sup>1,2</sup>, Matthew Hall<sup>1,2</sup>, George S. M. Dyer<sup>1,2</sup>, Brandon E. Earp<sup>1,2</sup> and Arvind von Keudell<sup>1,2</sup>

## Abstract

**Purpose:** Reverse total shoulder arthroplasty (rTSA) is effective and increasingly utilized for the management of proximal humerus fracture (PHF). However, the optimal patient-reported outcome metrics (PROMs) for the evaluation of patient outcomes after this surgery are unclear. We investigated the correlation among global, upper extremity-specific, and shoulder-specific PROMs in patients undergoing rTSA for PHF as well as the responsiveness of these PROMs as assessed by floor and ceiling effects. We hypothesized that patients' post-operative outcome would be best reflected by a combination of these metrics.

**Methods:** Thirty patients with a history of rTSA for ipsilateral PHF filled out the following outcomes questionnaires at a minimum of 3 years post-op: EQ-5D, EQ-5D VAS, PROMIS physical function, DASH, SSV, SPADI, and ASES. Correlation between metrics was assessed using the Spearman correlation coefficient. Responsiveness was assessed by comparing the proportion of patients reaching floor or ceiling values using McNemar's test.

**Results:** Global health metrics (EQ-5D and PROMIS physical function) were strongly correlated with the upper extremity-specific metric (DASH). Shoulder-specific outcomes (SPADI, ASES, and ASES) were moderately correlated with both the global metrics and DASH. There was no significant difference between PROMs with regards to floor and ceiling effects.

**Conclusions:** The DASH score has been shown to be valid and responsive for shoulder interventions, and our data demonstrate that it correlates strongly with overall quality of life. Shoulder-specific metrics are valid and responsive for shoulder interventions but correlate less with global quality of life. An optimal PROM strategy in rTSA for PHF might involve both DASH and a shoulder-specific score. Based on our assessment of floor and ceiling effects, none of these metrics should be excluded for poor responsiveness.

**Keywords:** Proximal humerus fracture, Reverse total shoulder arthroplasty, Functional outcome, Correlation, Responsiveness, Floor effect, Ceiling effect

\* Correspondence: [jbarger@mgh.harvard.edu](mailto:jbarger@mgh.harvard.edu)

<sup>1</sup>Department of Orthopaedic Surgery, Brigham and Women's Hospital, 55 Fruit Street, MA 02115 Boston, USA

<sup>2</sup>Harvard Medical School, 25 Shattuck St, 02115 Boston, MA, USA



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Introduction

Fractures of the proximal humerus account for nearly 6 % of fractures in the adult population and the frequency of this injury is expected to increase with aging of the US population [1]. Non-operative management is indicated in the majority of cases, [2, 3] but surgery may be considered for displaced, unstable 3- and 4-part fractures, head-split fractures, and fracture-dislocations [4]. While the optimal mode of surgical management of proximal humerus fractures (PHFs) continues to be debated, [4] the use of reverse total shoulder arthroplasty (rTSA) for the surgical treatment of complex PHFs is increasing [5–7]. Numerous studies have demonstrated the effectiveness of rTSA and better outcomes for rTSA than hemiarthroplasty, though no randomized study has demonstrated superior outcomes from rTSA relative to non-operative management [7–13]. Prior studies on rTSA for fracture have reported patient outcomes using a variety of global health, upper extremity-specific, and shoulder-specific functional scores [4, 14].

Patient-reported outcome metrics (PROMs) attempt to capture the biopsychosocial impact of disease on patients' lives [15] and are becoming more important than ever in orthopaedic surgery with the shift in healthcare from fee-for-service to value-based care [16]. They allow the patient's subjective experience of their injury and its treatment to be formally assessed, in addition to traditional objective and clinician-reported measures, which tend to underestimate patient symptoms and functional limitations [17, 18]. Health-related quality of life (HRQoL) is defined as how a person's health affects his or her ability to carry out normal social and physical activities [19]. Surgery for a PHF is intended to restore HRQoL and reduce disability, but the relationship between overall HRQoL and extremity-specific function for this condition is not known. In addition, PROMIS scores are increasingly used to evaluate patient functional outcomes but they have not been widely studied for upper extremity conditions [16]. Understanding associations between global health, upper extremity-specific, and shoulder-specific outcomes after rTSA for fracture will help surgeons to better measure the results of this treatment on patients' lives.

The objectives of this study were (1) to assess correlations among global health, upper extremity-specific, and shoulder-specific functional outcome measures for rTSA performed for proximal humerus fractures at minimum 3-year follow-up, and (2) to compare the responsiveness of global health, upper extremity-specific, and shoulder-specific functional outcome measures by assessing floor and ceiling effects.

## Methods

### Study design and patient selection

The study was designed following the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) guidelines for studies on measurement properties and specifically those guidelines regarding construct validity and responsiveness [20]. The study meets COSMIN criteria for these psychometric parameters, though a larger patient population (> 50) would have been preferred. Other psychometric parameters of the PROMs included in the study, including internal consistency and content, structural, and cross-cultural validity were outside the scope of this study but have been reported by previous investigators.

After obtaining approval from the Institutional Review Board, a retrospective analysis was performed of all patients who underwent rTSA for proximal humerus fractures at two American College of Surgeons Level I trauma centers from January 2003 to June 2017. Using our institutions' Research Patient Data Registry (RPDR), patients were identified using the International Classification of Disease (ICD-9 and ICD-10) codes for the diagnosis of proximal humerus fracture and Current Procedural Terminology (CPT) code for shoulder arthroplasty. Five hundred seventy-six patients were identified from the initial screen. Inclusion criteria were patients 18 years of age or older who underwent rTSA for a proximal humerus fracture between January 1, 2003 and August 1, 2018 and had a minimum of 3 years postoperative follow-up. Exclusion criteria included diagnoses other than proximal humerus fracture, procedures other than rTSA, surgeries performed at an outside institution, and patients without documented postoperative follow-up.

We excluded 264 patients for a diagnosis other than proximal humerus fracture, 74 patients who did not undergo rTSA, and 14 patients who had rTSA for fracture performed at an outside institution. The remaining 224 patients were contacted by telephone at a minimum of 3 years following rTSA for our outcome measures; 71 patients were further excluded for inadequate follow-up, of which 26 were deceased and 55 were unavailable. Furthermore, 113 patients were excluded for an incomplete set of outcome measures at a minimum of 3 years following rTSA, 112 because they were not sent all of the functional outcomes surveys due to their inclusion in different research protocols, and 1 who was contacted for all scores but returned EQ-5D, PROMIS, and DASH but not ASES, SPADI, or SSV. The final cohort included 30 patients (30 shoulders) with proximal humerus fractures treated with rTSA with a complete set of functional outcome measures at least 3 years of follow-up.

## Outcome measures

The outcome measures in this study included global health outcome measures, an upper extremity-specific outcome measure, and shoulder-specific outcome measures. Global health outcome measures included the Patient-Reported Outcomes Measurement Information System (PROMIS) Global Health Physical Score Short Form 10b, EuroQol Five-Dimension Score (EQ-5D), and EuroQol Visual Analogue Scale (EQ-VAS). The upper extremity-specific outcome measure was the Disabilities of the Arm, Shoulder and Hand (DASH) score. Shoulder-specific outcome measures included the Shoulder Subjective Value (SSV), American Shoulder and Elbow Surgeons (ASES), and Shoulder Pain and Disability Index (SPADI). All outcome measures were collected through review of the medical records and telephone contact at a minimum of 3 years following rTSA.

The PROMIS Physical Function instrument is a 10-item questionnaire that evaluates physical, mental, and social aspects of health, and all items are evaluated on a 5-point Likert scale [10]. The questionnaire has Physical and Mental Health subscales. The Physical Health subscale score, which was utilized in our study, ranges from 0 to 100 with a mean score of 50 representing the norm for the US general population and a standard deviation of 10. Higher scores indicate better health [16, 21].

The EQ-5D is a health-related quality of life measure with 5 domains: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression [21]. There are three options for each response. A score of 1.0 represents perfect health while scores less than zero indicate health status worse than death [21]. The EQ-VAS is a separate quantitative self-assessment of health which has patients evaluate their own health on a visual analogue scale from 0 to 100 where 0 is worst and 100 is best [21].

The DASH score is a 30-item upper-extremity-specific outcome measure which is intended to measure shoulder, elbow, hand, and wrist function in one combined metric [22]. It does not discriminate between the affected and non-affected upper extremity. The DASH assesses multiple domains including physical function, symptoms, and social/psychological function [13]. It has been shown to correlate with shoulder-specific measures, [23]. The DASH is scored from 0 to 100, with lower scores representing better function. Normative values are 22 in females and 13 in males aged 70 to 79 [24]. The minimum clinically important difference (MCID) for shoulder conditions is approximately 10 [this MCID value, and those reported for other metrics below, is not intended to suggest that the MCID in our patient population is identical, but merely to give a sense of the scale on which clinically important differences have been previously reported to occur for other shoulder pathologies] [22].

The SSV is an entirely subjective shoulder-specific measure where the patient is asked to rate their shoulder on a scale from 0 to 100 with some version of the following question: "What is the overall percent value of your shoulder if a completely normal shoulder represents 100%?" [25]. The SSV is considered an adjunct to more complicated shoulder-specific scores and has been shown to be an easily administered, responsive, and valid measure of shoulder function [25].

The ASES is a shoulder-specific score which consists of 10 functional questions and a pain VAS score [22]. The total score ranges from 0 to 100 (higher is better) and is weighted 50% for function and 50% for pain. The MCID is considered to be 6.4 for various shoulder pathologies [22].

The SPADI is a self-assessment of symptoms and function of the shoulder designed for any disorder of the shoulder joint [26]. It consists of 13 items (5 for pain symptoms and 8 for disability) which are scored on a 0 to 100 point visual response scale. The domain scores are combined to produce a score from 0 to 100 where 0 is best and 100 is worst [27]. The MCID for the SPADI has been reported to be between 8 and 13.2 [27].

Patient demographics were collected through our institutional RPDR and review of the electronic medical records, including age, sex, hand dominance, body mass index (BMI), Charlson Comorbidity Index (CCI), diabetes mellitus, depression, osteoporosis, smoking status at presentation, and employment status. Injury-related variables included fracture type, presence of ipsilateral upper extremity fracture, open fracture, and associated nerve injury. Treatment-related variables included initial treatment (acute rTSA versus nonoperative treatment), tuberosity repair, complications, and reoperation. The operative techniques and rehabilitation protocols were at the discretion of the treating surgeon.

## Statistical analysis

Descriptive statistics were calculated for the study cohort. Measurement of correlations between outcome measures were performed using the Spearman correlation coefficient. To study floor and ceiling effects, the proportion of patients reporting the minimum and maximum scores of each outcome measure, respectively, were determined. Paired comparisons of the proportion of patients at the floor and ceiling of outcome measures were performed using McNemar's test. The standard significance criterion of  $\alpha = 0.05$  was employed. A convenience sample was used.

## Results

### Cohort demographics

Our study cohort comprised 30 patients treated with rTSA following proximal humerus fracture. The average

age of our cohort was 70.2 years at the time of surgery, and 87% were female. Ten (33%) of patients had diabetes mellitus, 6 (20%) had depression, 6 (20%) had osteoporosis, and 1 (3%) smoked at the time of surgery. The dominant upper extremity was affected in 18 (60%) of patients, and 6 (20%) of patients were actively employed. The tuberosity was repaired during rTSA in 20 (67%) of patients (Table 1).

Twenty-three out of 30 (77%) patients were treated with rTSA acutely, and 7 (23%) were treated with rTSA after failed initial nonoperative treatment. Complications occurred in four patients and included instability, infection, heterotopic ossification, and periprosthetic fracture. Two patients (7%) underwent reoperation, which included component revision for instability and irrigation and debridement with liner exchange for infection. The median clinical follow-up after rTSA was 17.5 months (IQR 10.5 to 30.1 months). The median time to telephone assessment of functional outcome measures was 5.2 years (IQR 4.3 to 6.7 years, minimum 59 days).

#### Correlation between outcome measures

Global health, upper extremity-specific outcome measure, and shoulder-specific outcome measures were collected for the cohort at minimum 3-year follow-up after rTSA for fracture (Table 2).

**Table 1** Baseline characteristics of rTSA for fracture patients ( $n = 30$ )

	Mean (Standard Deviation)
Age	70.2 (6.7)
BMI	30.1 (5.1)
	Median (Interquartile Range)
CCI	3 (2–5)
	n (%)
Male sex	4 (13)
Diabetes mellitus	10 (33)
Depression	6 (20)
Osteoporosis	6 (20)
Current smoker	1 (3)
Employed	6 (20)
Dominant upper extremity injury	18 (60)
Ipsilateral upper extremity injury	1 (3)
Nerve injury	1 (3)
Cemented humeral stem	14 (47)
Tuberosity repaired	20 (67)

**Table 2** Functional outcome scores of rTSA for fracture patients ( $n = 30$ )

	Mean (Standard Deviation)
PROMIS	43.7 (9.9)
EQ-5D	0.77 (0.13)
EQ-VAS	76.5 (13.6)
DASH	25.2 (16.7)
SSV	78.4 (21.5)
SPADI	26.4 (18.8)
ASES	73.6 (20.7)

Global health and upper extremity-specific outcome measures were compared (Fig. 1). PROMIS and DASH ( $\rho = -0.85$ ,  $p < 0.0001$ ) and EQ-5D and DASH ( $\rho = -0.75$ ,  $p < 0.0001$ ) were highly correlated. EQ-VAS and DASH ( $\rho = -0.45$ ,  $p = 0.002$ ) were moderately correlated.

Global health and shoulder-specific outcome measures were compared (Fig. 2). PROMIS and SSV ( $\rho = 0.58$ ,  $p = 0.0007$ ), PROMIS and SPADI ( $\rho = -0.55$ ,  $p = 0.002$ ), and PROMIS and ASES ( $\rho = 0.51$ ,  $p = 0.004$ ) were moderately correlated. EQ-5D and SSV ( $\rho = 0.63$ ,  $p = 0.0002$ ), EQ-5D and SPADI ( $\rho = -0.47$ ,  $p = 0.008$ ), and EQ-5D and ASES ( $\rho = 0.48$ ,  $p = 0.007$ ) were moderately correlated. EQ-VAS and SSV ( $\rho = 0.55$ ,  $p = 0.002$ ) were moderately correlated, but EQ-VAS and SPADI ( $\rho = -0.35$ ,  $p = 0.06$ ) and EQ-VAS and ASES ( $\rho = 0.29$ ,  $p = 0.1$ ) were not significantly correlated.

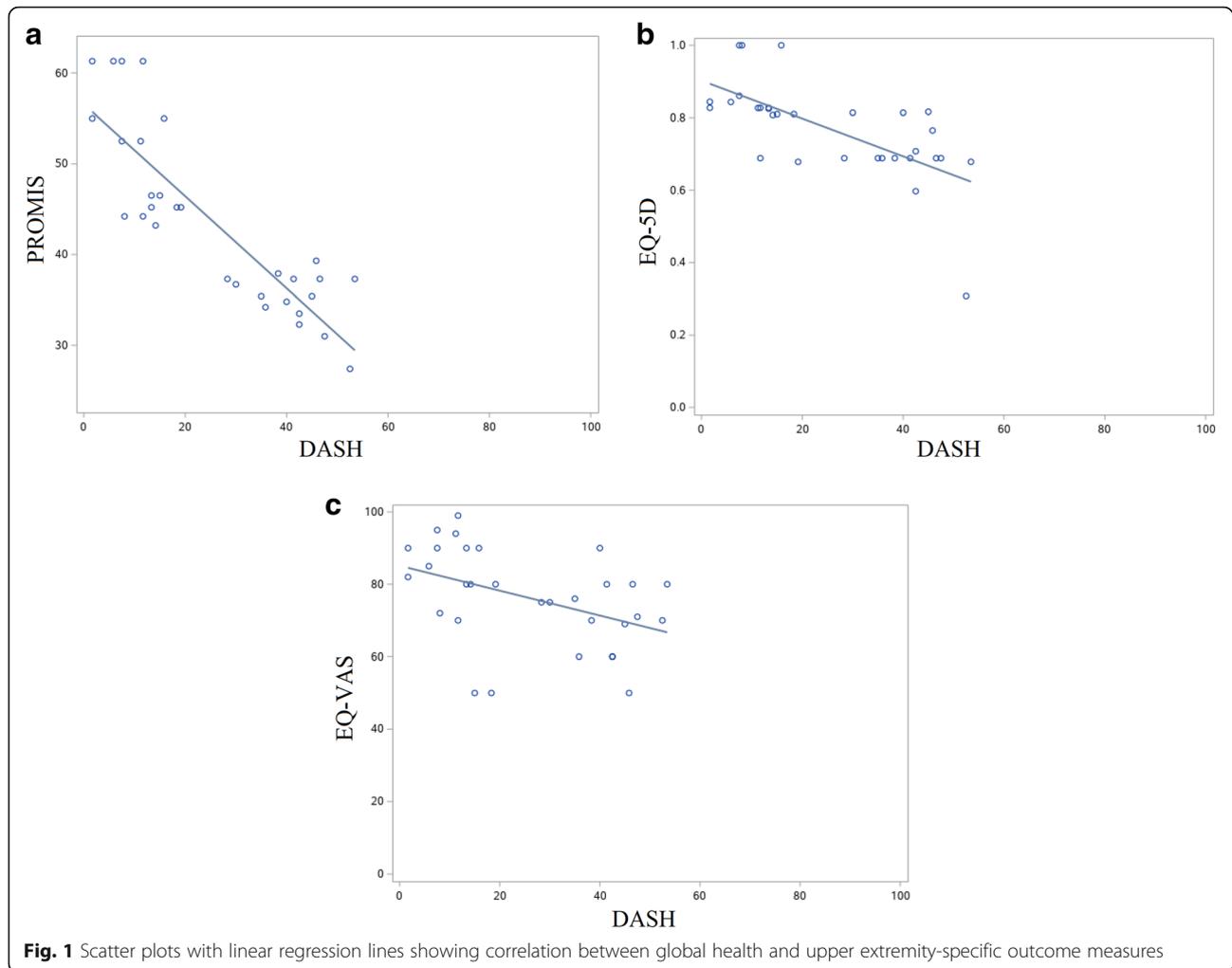
Upper extremity-specific and shoulder-specific outcome measures were compared (Fig. 3). DASH and SSV ( $\rho = -0.67$ ,  $p < 0.0001$ ), DASH and SPADI ( $\rho = 0.55$ ,  $p = 0.002$ ), and DASH and ASES ( $\rho = -0.52$ ,  $p = 0.003$ ) were moderately correlated.

Among global health outcome measures, PROMIS and EQ-5D ( $\rho = 0.73$ ,  $p < 0.0001$ ) were strongly correlated, whereas PROMIS and EQ-VAS ( $\rho = 0.59$ ,  $p = 0.0005$ ) and EQ-5D and EQ-VAS ( $\rho = 0.54$ ,  $p = 0.002$ ) were moderately correlated. Among shoulder-specific outcome measures, SSV and SPADI ( $\rho = -0.67$ ,  $p < 0.0001$ ), SPADI and ASES ( $\rho = -0.68$ ,  $p < 0.0001$ ), and SSV and ASES ( $\rho = 0.69$ ,  $p < 0.0001$ ) were moderately correlated.

#### Responsiveness of outcome measures

Five patients reported the maximum SSV score, 3 patients reported the maximum EQ-5D score, and 1 patient reported the maximum SPADI score. No other functional outcome measures exhibited a ceiling effect.

One patient reported the minimum ASES score. No other functional outcome measure exhibited a floor effect.



**Fig. 1** Scatter plots with linear regression lines showing correlation between global health and upper extremity-specific outcome measures

McNemar's tests did not demonstrate a significant difference between the proportion of patients at the ceilings of EQ-5D and SSV ( $p = 0.4$ ) and EQ-5D and SPADI ( $p = 0.3$ ).

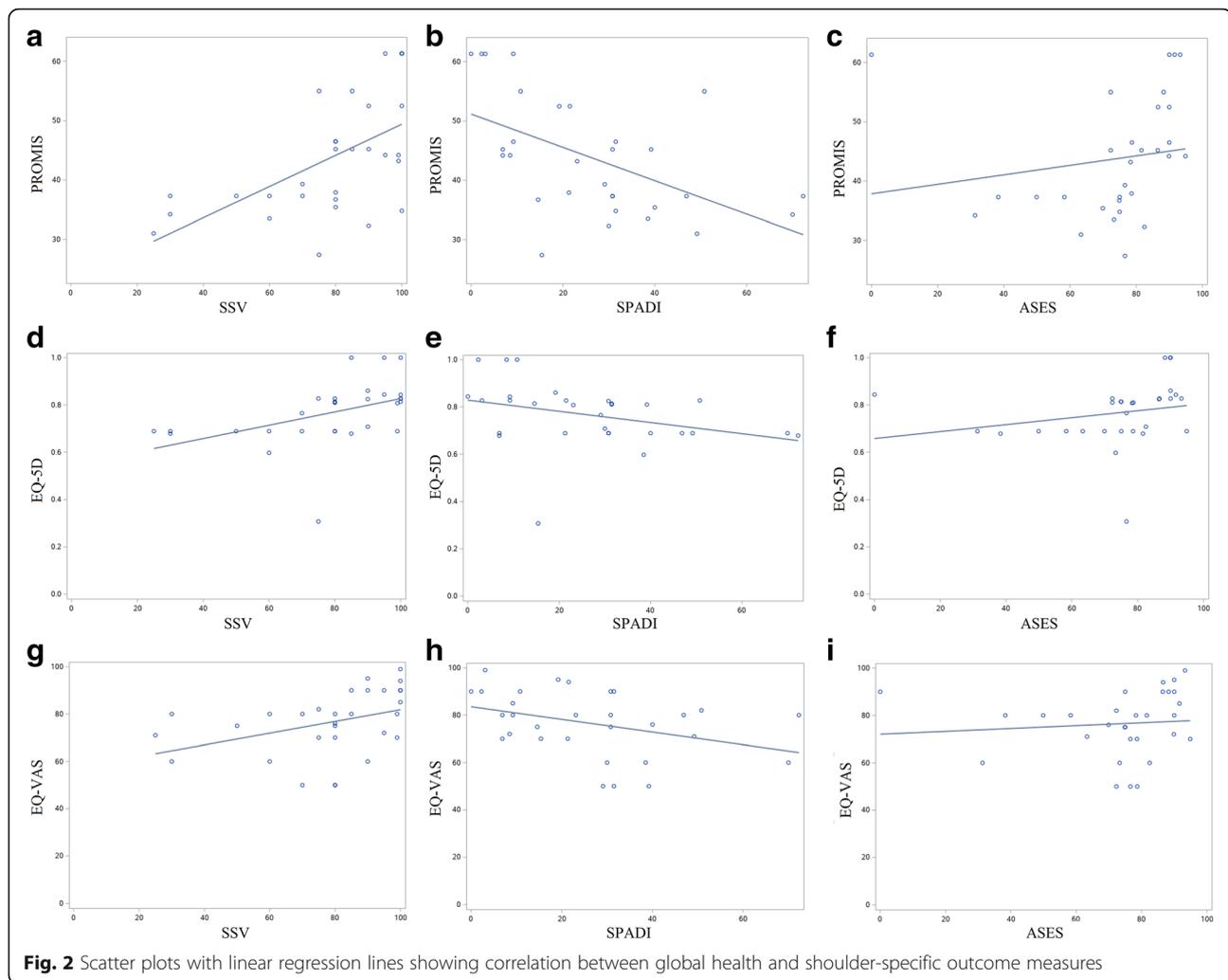
## Discussion

There are numerous PROMs available for the assessment of HRQoL and function following shoulder surgery. In selecting which measures to collect from patients, it is important to understand which aspects of the patient's experience are reflected by each score and to avoid unnecessary redundancy [28]. Rates of survey completion by patients has been shown to correlate negatively with survey length, suggesting that it is important to select outcome measures judiciously [28]. PROMs collected in previous studies on rTSA for fracture have varied widely, so our study sought to investigate the contribution of and relationship between several commonly used metrics [4, 14].

The correlation between global outcome measures (EQ-5D and PROMIS physical function) and DASH was stronger than that between global outcomes and

shoulder-specific measures (ASES, SPADI, and SSV). This likely reflects the fact that the DASH is a comprehensive evaluation of upper-extremity function that captures the effect of an upper-extremity intervention (rTSA for fracture, in this case) on a patient's overall quality of life. As such, it can be interpreted as a proxy for the effect of an upper extremity intervention on global HRQoL, while simultaneously focusing on the upper extremity and thereby being more sensitive to treatment effect than truly global metrics like the EQ-5D. This hypothesized increased sensitivity is borne out by the fact that the data in this study was used in a separate project, not yet published, to compare outcomes between those undergoing rTSA for fracture acutely versus in a delayed manner for mal- or non-union. While there was a significant difference in DASH scores; EQ-5D and PROMIS did not detect this group difference.

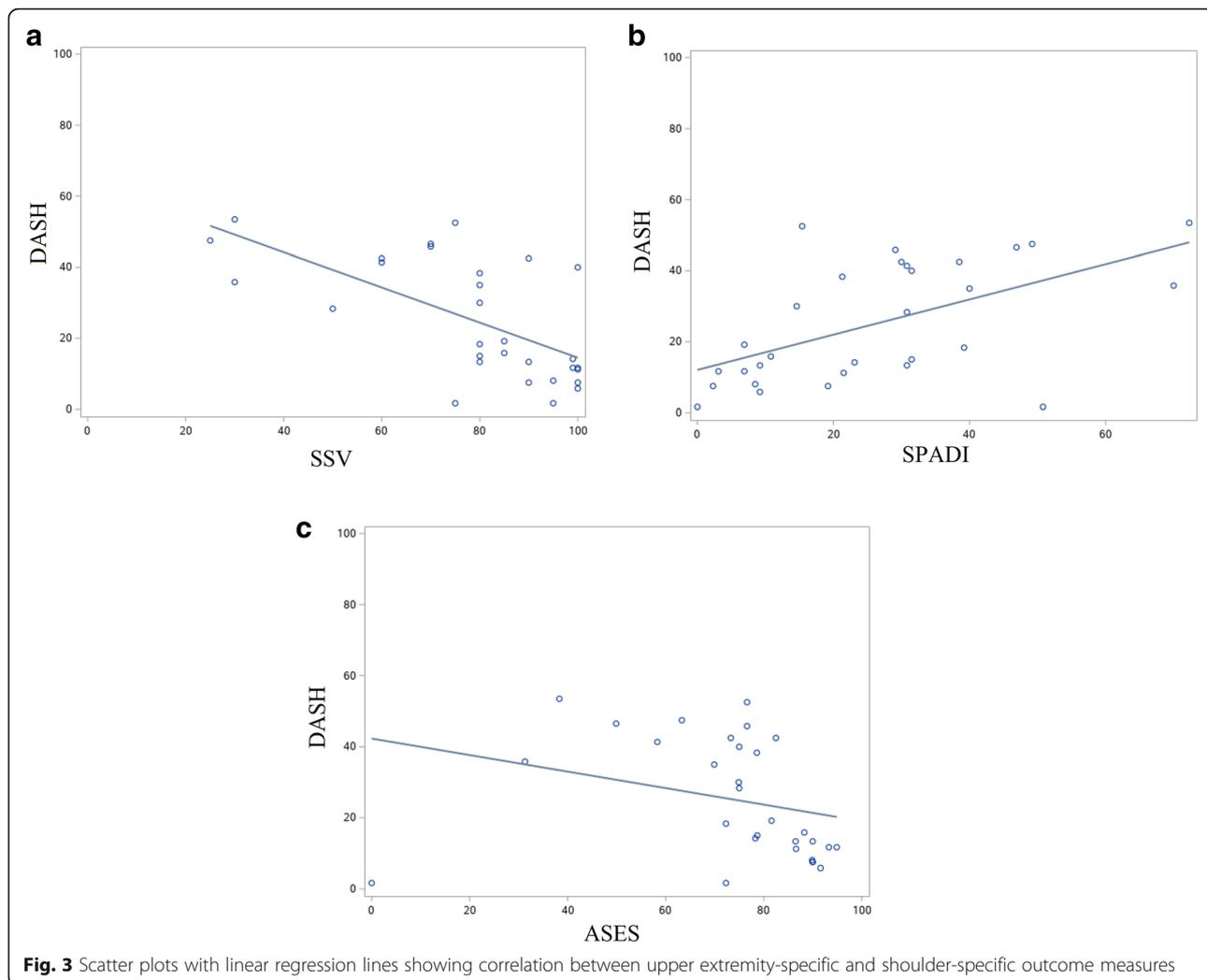
Shoulder-specific outcomes scores (ASES, SPADI, and SSV) correlated only moderately with DASH scores and global metrics. These scores focus explicitly on shoulder function, and while they may be responsive to shoulder



interventions, are less reflective of overall quality of life. These findings suggest that the optimal selection of outcomes scores in evaluating rTSA for fracture may include DASH score, which offers upper-extremity focus while correlating well with global HRQoL, plus a shoulder-specific score to optimize responsiveness to this shoulder intervention. Our findings also support the use of the PROMIS instrument as an easy-to-use alternative to the well-established but lengthier DASH questionnaire. Since we found significant correlation among shoulder-specific outcome measures (ASES, SPADI, and SSV), we recommend collecting only one shoulder-specific score to avoid redundancy. Our study supports the construct validity of the Subjective Shoulder Score as an easily obtained metric that reflects patient quality of life as it relates to the shoulder. The SSV correlated more closely (though still just moderately) with DASH, PROMIS, and EQ-5D than ASES or SPADI and involves only one simple question. Therefore, for researchers who feel collecting the DASH metric in addition to a

shoulder-specific score may add undue burden, the SSV might be a reasonable alternative.

While collecting both DASH and a shoulder-specific metric seems to be a good combination following rTSA for fracture, this raises the question of whether two PROMs may introduce survey fatigue, particularly as the DASH contains 30 questions and is therefore one of the longer survey metrics available. A meta-analysis of survey burden in PROMs by Rolstad et al. demonstrated that increased survey length correlates with decreased response rate. However, there was no particular survey length at which responses dropped off and the willingness of patients to complete longer surveys varied between conditions [28]. As such, there is no clear answer to how much is “too much.” Atkinson et al. found minimal patient-reported survey burden in surgical oncology patients completing the 30-question EORTC QLQ-C30 metric, whether or not it was coupled with a 30-minute interview on their experiences of illness [29]. This was despite the fact that many subjects noted redundancy



and suggests that patients are generally willing to spend considerable time reporting on conditions which have greatly affected their lives. As such, we feel that collecting two outcome metrics is unlikely to decrease survey response, but cannot conclude this definitively. This decision ultimately resides with clinician-researchers. In our study, of the 31 patients who were contacted for all six PROMs and agreed to participate, just one failed to complete them all, suggesting that the survey burden of two PROMs is unlikely to be an obstacle to data collection in patients undergoing rTSA for fracture.

In addition to investigating the correlation between outcomes scores in patients undergoing rTSA for fracture, our study sought to assess whether the responsiveness of these scores were affected by floor or ceiling effects. The floor or ceiling effect describes the circumstance in which patient outcome scores cluster at the minimum or maximum value for a given metric, thus rendering the outcome measure unable to detect differences between these patients [30]. At least one patient

reported the maximum value for three of the metrics (SSV, EQ-5D, and SPADI), but for each of these metrics the number of patients at the ceiling was small, and we found no significant difference in the proportion of patients at the ceiling. It is unsurprising we did not find a ceiling effect, given that these patients did not necessarily have any pathology prior to their fracture and thus their pre-injury comparison was likely to a fairly normal shoulder. More interesting is the fact that only one patient reported a minimum value, for the ASES. These findings suggest that the responsiveness of the outcomes we measured is not affected by floor or ceiling effects in rTSA for fracture.

There are numerous additional outcomes metrics available for the study of shoulder interventions which were not included in our study. Perhaps the most widely used instrument in the proximal humerus fracture literature is the Constant-Murley score. We elected not to collect this score and feel it is not the metric of choice in studies of rTSA after fracture for multiple reasons.

First, it contains a physical exam component which includes range of motion and power. This introduces the risk of bias by the examiner, may be inappropriate to combine with patient-reported domains, and is conflated by normal age-related declines in shoulder strength [31]. Second, several studies have questioned its reliability and correlation with other shoulder outcomes metrics [31]. Two other metrics which were not included in our study but which are widely used in shoulder arthroplasty registries are the Western Ontario Osteoarthritis of the Shoulder (WOOS) index (particularly used in Scandinavian registries) and the Oxford Shoulder Score (OSS) (widely used in Commonwealth registries). The WOOS is a disease-specific metric designed to be particularly responsive to interventions for shoulder osteoarthritis [32]. The Pearson correlation coefficient between the WOOS and ASES has been reported between 0.570 in the original validation study to 0.88 more recently in patients undergoing rTSA for OA, representing a moderate to excellent correlation [32, 33]. There is also strong correlation between the WOOS and DASH ( $r = 0.75$ ) [34]. These data suggest that there is translatability between the WOOS and the metrics in our study. The OSS has been reported to have excellent correlation with the ASES ( $r = 0.91$ ) [35] and moderate correlation with DASH ( $r = -0.059$ ) [36], suggesting translatability of the OSS as well.

Limitations of this study include the fact that several widely utilized PROMs were not included due to concern for patient survey fatigue. We did not collect the questionnaires at multiple pre-determined time points and thus were unable to compare the proportion of patients reaching clinically important differences, which would have better allowed us to assess the responsiveness of the scores over time. Finally, the patient population was on the smaller side of the acceptable range recommended by COSMIN guidelines.

PROMs allow the patient's subjective experience to be reflected in surgical outcomes, but there has been little consistency in the literature as to which PROMs should be used to track outcomes after rTSA for fracture. Our results suggest that use of a global health or upper extremity-specific score, such as PROMIS physical function or DASH, in conjunction with a shoulder-specific metric, such as SSV, SPADI, or ASES, may strike an optimal balance. The PROMIS physical function score and SSV are appealing instruments, because they correlate well with other relevant metrics and are more easily administered. There was no significant difference in floor or ceiling effects between any of the metrics in our study.

#### Acknowledgements

none.

#### Authors' contributions

JB, DZ, and DS wrote the main manuscript text. DZ analyzed data and prepared Figs. 1, 2 and 3. JB, DS, PO, and MH were involved in data gathering and analysis. GD and BE provided senior oversight and guidance. DZ and AvK provided senior oversight and conceived study design. The author(s) read and approved the final manuscript.

#### Funding

No sources of funding to disclose.

#### Availability of data and materials

De-identified data may be requested from the corresponding author via email.

#### Declarations

#### Ethics approval and consent to participate

This study was performed with institutional review board approval and all methods were carried out in accordance with relevant guidelines and regulations. Informed consent was obtained from all subjects for use and publication of their de-identified data in IRB-approved research. Approval was provided by the Mass General Brigham Institutional Review Board.

#### Consent for publication

This study was performed with institutional review board approval and all methods were carried out in accordance with relevant guidelines and regulations. Informed consent was obtained from all subjects for use and publication of their de-identified data in IRB-approved research. Approval was provided by the Mass General Brigham Institutional Review Board.

#### Competing interests

The authors have no financial conflict or sources of financial or material funding to disclose. No portion of this work has been previously presented or published.

Received: 25 January 2021 Accepted: 1 June 2021

Published online: 23 June 2021

#### References

1. Court-Brown CM, Caesar B. Epidemiology of adult fractures: A review. *Injury*. 2006;37(8):691–7.
2. Rangan A, Handoll H, Brealey S, Jefferson L, Keding A, Martin B, et al. Surgical vs nonsurgical treatment of adults with displaced fractures of the proximal humerus: the PROFHER randomized clinical trial. *JAMA*. 2015; 313(10):1037–47.
3. Orman S, Mohamadi A, Serino J, Murphy J, Hanna P, Weaver MJ, et al. Comparison of surgical and non-surgical treatments for 3- and 4-part proximal humerus fractures: A network meta-analysis. *Shoulder Elbow*. 2020; 12(2):99–108.
4. Gupta AK, Harris JD, Erickson BJ, Abrams GD, Bruce B, McCormick F, et al. Surgical Management of Complex Proximal Humerus Fractures—A Systematic Review of 92 Studies Including 4500 Patients. *Journal of Orthopaedic Trauma*. 2015;29(1):54–9.
5. Kim S, Wise B, Zhang Y, Szabo R. Increasing incidence of shoulder arthroplasty in the United States. *J Bone Joint Surg Am*. 2011;93(24): 2249–54.
6. Acevedo DC, Mann T, Abboud JA, Getz C, Baumhauer JF, Voloshin I. Reverse total shoulder arthroplasty for the treatment of proximal humeral fractures: patterns of use among newly trained orthopedic surgeons. *Journal of Shoulder and Elbow Surgery*. 2014;23(9):1363–7.
7. Acevedo D, Vanbeek C, Lazarus M, Williams G, Abboud J. Reverse shoulder arthroplasty for proximal humeral fractures: update on indications, technique, and results. *J Shoulder Elbow Surg*. 2014;23(2).
8. Chalmers P.N. SW, Mall N.A., Gupta A.K., Rahman Z., Enriquez D., Nicholson G. Reverse total shoulder arthroplasty for acute proximal humeral fracture: open reduction-internal fixation and hemiarthroplasty. *J Shoulder Elbow Surg*. 2014;23:197–204.
9. Lopiz Y, Alcobia-Díaz B, Galán-Olleros M, García-Fernández C, Picado AL, Marco F. Reverse Shoulder Arthroplasty Versus Nonoperative Treatment for 3- Or 4-part Proximal Humeral Fractures in Elderly Patients: A Prospective Randomized Controlled Trial. *J Shoulder Elbow Surg*. 2019;28(12):2259–71.

10. Gallinet D, CP, Garbuio P, Tropet Y, and Obert L. Three or four parts complex proximal humerus fractures: hemiarthroplasty versus reverse prosthesis: a comparative study of 40 cases. *Orthop Traumatol Surg Res.* 2009;95:48–55.
11. Klein M, JM, Hinkenjann B, Scherger B, and Ostermann P.A. Treatment of comminuted fractures of the proximal humerus in elderly patients with the Delta III reverse shoulder prosthesis. *J Orthop Trauma.* 2008;22:698–704.
12. Dezfuli B, King, JJ, Farmer, KW, Struk, AM, Wright, TW. Outcomes of reverse total shoulder arthroplasty as primary versus revision procedure for proximal humerus fractures. *J Shoulder Elbow Surg.* 2016;25(7):1133–7.
13. Spross C, Meester J, Mazzucchelli R, Puskás G, Zdravkovic V, Jost B. Evidence-based algorithm to treat patients with proximal humerus fractures—a prospective study with early clinical and overall performance results. *J Shoulder Elbow Surg.* 2019;28(6):1022–32.
14. Gallinet D, Cazeneuve J-F, Boyer E, Menu G, Obert L, Ohl X, et al. Reverse shoulder arthroplasty for recent proximal humerus fractures: Outcomes in 422 cases. *Orthopaedics & Traumatology: Surgery & Research.* 2019;105(5): 805–11.
15. Jayakumar P, Williams M, Ring D, Lamb S, Gwilym S. A Systematic Review of Outcome Measures Assessing Disability Following Upper Extremity Trauma. *J Am Acad Orthop Surg Glob Res Rev.* 2017;1(4):e021-e.
16. Stoop N, Menendez ME, Mellema JJ, Ring D. The PROMIS Global Health Questionnaire Correlates With the QuickDASH in Patients With Upper Extremity Illness. *Hand (N Y).* 2018;13(1):118–21.
17. Deshpande PR, Rajan S, Sudeepthi BL, Abdul Nazir CP. Patient-reported outcomes: A new era in clinical research. *Perspect Clin Res.* 2011;2(4): 137–44.
18. Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. *J Natl Cancer Inst.* 2009;101(23):1624–32.
19. Poolman RW, Swiontkowski MF, Fairbank JCT, Schemitsch EH, Sprague S, de Vet HCW. Outcome instruments: rationale for their use. *J Bone Joint Surg Am.* 2009;91 Suppl 3(Suppl 3):41–9.
20. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology.* 2010;10(1):22.
21. Shim J, Hamilton DF. Comparative responsiveness of the PROMIS-10 Global Health and EQ-5D questionnaires in patients undergoing total knee arthroplasty. *Bone Joint J.* 2019;101-B(7):832–7.
22. Wylie JD, Beckmann JT, Granger E, Tashjian RZ. Functional outcomes assessment in shoulder surgery. *World J Orthop.* 2014;5(5):623–33.
23. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? *Journal of Hand Therapy.* 2001;14(2):128–42.
24. Aasheim T, Finsen V. The DASH and the QuickDASH instruments. Normative values in the general population in Norway. *Journal of Hand Surgery (European Volume).* 2013;39(2):140–4.
25. Gilbert MK, Gerber C. Comparison of the subjective shoulder value and the Constant score. *Journal of Shoulder and Elbow Surgery.* 2007;16(6):717–21.
26. Angst F, Schwyzer H-K, Aeschlimann A, Simmen BR, Goldhahn J. Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and Its Short Version (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society Standardized Shoulder. *Arthritis Care & Research.* 2011;63(S11):S174–S88.
27. Roy J-S, MacDermid JC, Woodhouse LJ. Measuring shoulder function: A systematic review of four questionnaires. *Arthritis & Rheumatism.* 2009;61(5): 623–32.
28. Rolstad S, Adler J, Rydén A. Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-analysis. *Value in Health.* 2011;14(8): 1101–8.
29. Atkinson TM, Schwartz CE, Goldstein L, Garcia I, Storfer DF, Li Y, et al. Perceptions of Response Burden Associated with Completion of Patient-Reported Outcome Assessments in Oncology. *Value in Health.* 2019;22(2): 225–30.
30. Broughton V. *Sage Dictionary of Statistics: A Practical Resource for Students in the Social Sciences* 200585Duncan Cramer and Dennis Howitt. Sage Dictionary of Statistics: A Practical Resource for Students in the Social Sciences. London and Thousand Oaks, CA: Sage 2004. x + 188 pp., ISBN: 0 7619 4137 1 (hbk); 0 7619 4138 X (pbck) £60 \$94.95 (hbk); £18.99 \$37.95 (pbck). *Reference Reviews.* 2005;19(2):26–7.
31. Wright RW, Baumgarten KM. Shoulder outcomes measures. *J Am Acad Orthop Surg.* 2010;18(7):436–44.
32. Lo IKY, Griffin S, Kirkley A. The development of a disease-specific quality of life measurement tool for osteoarthritis of the shoulder: The Western Ontario Osteoarthritis of the Shoulder (WOOS) index. *Osteoarthritis and Cartilage.* 2001;9(8):771–8.
33. Baumgarten KM. The American Shoulder and Elbow Surgeons Score has Excellent Correlation with the Western Ontario Osteoarthritis Score. *J Shoulder Elbow Surg.* 2021.
34. Corona K, Cerciello S, Morris BJ, Visonà E, Merolla G, Porcellini G. Cross-cultural adaptation and validation of the Italian version of the Western Ontario Osteoarthritis of the Shoulder index (WOOS). *J Orthop Traumatol.* 2016;17(4):309–13.
35. Haparachchi KS, Poon PC. A correlation study of the American Shoulder and Elbow Society Score and the Oxford Shoulder Score with the use of regression analysis to predict one score from the other in patients undergoing reverse shoulder joint arthroplasty for cuff tear arthropathy. *Shoulder Elbow.* 2014;6(2):81–9.
36. Ebrahimzadeh MHM, Birjandinejad AM, Razi SM, Mardani-Kivi MM, Reza Kachooei AM. Oxford Shoulder Score: A Cross-Cultural Adaptation and Validation Study of the Persian Version in Iran. *Iran J Med Sci.* 2015;40(5): 404–10.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

