JSES International 7 (2023) 653–661

Contents lists available at ScienceDirect

JSES International

journal homepage: www.jsesinternational.org

Trends in shoulder surgery patient-reported outcome measures

Rajpal Narulla, MBBS, MS[a,b,*,1], Mark Song, MBBS, MS[b], Sascha Karunaratne, BHlthSci(Hon)/MPhty[c], Christopher Smithers, MBBS, FRACS, FAOrthA[a], Jeffrey Petchell, MBBS, MMed (Clin Epi), FRACS, FAOrthA[a]

[a]*Department of Orthopaedic Surgery, Royal Prince Alfred Hospital, Camperdown, NSW, Australia*
[b]*University of New South Wales, Sydney, NSW, Australia*
[c]*Surgical Outcomes Research Centre (SOuRCe), Royal Prince Alfred Hospital, Camperdown, NSW, Australia*

ARTICLE INFO

Keywords:
PROMs
Shoulder surgery
Arthroplasty
Instability
Rotator cuff
Patient reported outcome measures

Level of evidence: Systematic Review; Literature Review of Outcome Instruments

Patient-reported outcome measures (PROMs) are a vital part of the toolkit for the current practice of orthopedic surgery. We are witnessing the expansion of the use of PROMs in clinical practice and in research; the ultimate direction of this expansion is unclear. The purpose of this systematic review was to identify the trends in the use of PROMs in major upper limb publications over a 7-year period. We retrospectively reviewed all articles published in 6 of the most influential upper limb orthopedic journals based on impact factor from January 2013 to January 2020. PubMed, Medline, and Embase were used to access the abstracts for all articles published for this period. We included all articles related to shoulder arthroplasty, shoulder instability, rotator cuff surgery, and involving the use of PROMs. There were 4175 articles identified from the selected journals over the chosen time period, of which 607 were eligible for inclusion in the study. The number of articles reporting PROMs increased from 57 in 2013 to 115 in 2019, which was a 102% increase. The total number of PROM usages recorded was 1593 which was comprised of 63 different scoring systems, with each article using a median of 3 different PROMs. The most commonly used score in articles originating from North America was the American Shoulder and Elbow Surgeons score (216 uses in 273 articles; 78.1%), from Europe it was the Constant-Murley Score (129 uses in 183 articles; 70.4%), and from Asia it was the American Shoulder and Elbow Surgeons score (80 uses in 126 articles; 63.4%). The use of PROMs is evolving with an increasing prevalence of and diversity of PROMs being used in upper limb surgery. There is geographical variation in the use of PROMs, and a variety of systems used, with only 3 of the top 10 most used PROMs reporting on patient satisfaction or wellbeing. Given that a diverse range of PROMs study a diverse range of conditions and procedures, there may not be a need for a consensus on the best overall use of PROMs, but there may be ideal PROMs suited to answer specific questions.

As the standard approach to evaluating the success of surgery pivots toward patient-centered care in orthopedic surgery, clinicians are faced with an abundance of tools to evaluate different facets of their procedures. In recent years, patient-reported outcome measures (PROMs) have become a key consideration in understanding the effects of surgical procedures. They have been used with great success to evaluate the effectiveness of operations and provide clinicians and patients alike with an understanding of what they can expect after surgery.[1,13] PROMs are particularly important in shoulder surgery where the difference between a good and a bad result for a patient may depend on the ability to achieve a functional goal, or a feeling of stability, which cannot be quantified by radiographic markers of operative success.[21,24] Evidence-based clinicians must consider how they can best compare their own results against their colleagues' results and use tools acceptable to peer-reviewers when attempting to publish their results.[15] While there are benefits to having an unlimited

Institutional review board approval/ethical review was not sought as this is a Systematic Review of published literature which does not require institutional review board approval at the authors' institution(s).

This work was performed at the Department of Orthopaedic Surgery, Royal Prince Alfred Hospital, Level 10, West 2, 50 Missenden Road, Sydney, NSW, Australia 2050.

*Corresponding author: Rajpal Narulla, MBBS, MS, Department of Orthopaedic Surgery, Royal Prince Alfred Hospital, Level 10, West 2, 50 Missenden Road, Sydney, NSW 2050, Australia.

E-mail address: rsnarulla@gmail.com (R. Narulla).

[1] This author is supported by an Australian Government Research Training Program Scholarship.

https://doi.org/10.1016/j.jseint.2023.03.010

number of PROMs to choose from, validation of many of the less commonly used PROMs is lacking. This comment is applicable to PROMs that duplicate measures for common shoulder issues; however, there is a definite role for less commonly used PROMs that measure less common presentations such as tumor pathology. Another issue with the abundance of PROMs available is that many are interchangeable or overlapping in what they measure, which creates duplication and inefficiency in reporting.[14,17] Having standardization and validation of the PROMs we use is a step toward the elusive goals of understanding patient satisfaction and being able to predict what factors will lead to successful surgery.[16] Given the importance of sharing knowledge and standardizing results, we sought to evaluate the trend of PROMs usage for shoulder surgery studies published in major journals in recent years. The aim of this review is to show which PROMs are most prevalent in the literature and highlight differences in PROMs usage in different time periods, by different journals, and by different regions of publication.

Materials and methods

We retrospectively reviewed all articles published in 6 of the most influential upper limb orthopedic journals based on impact factor from January 2013 to January 2020. We used PubMed, Medline, and Embase to access the abstracts for all articles published for this period in the American Journal of Sports Medicine (IF: 6.057), the Journal of Bone and Joint Surgery (U.S.A; IF: 4.578), Arthroscopy (IF: 4.325), the Bone and Joint Journal (UK; IF 4.306), Clinical Orthopaedics and Related Research (IF: 4.091), and the Journal of Shoulder and Elbow Surgery (IF: 2.817). These journals were selected based on impact factor and relevance to the field of shoulder surgery. We screened all articles published during this period to exclude duplicates and include those related to shoulder arthroplasty, shoulder instability, or rotator cuff surgery and involving the use of PROMs in the assessment of patients. These subtypes of surgery were chosen as they cover most orthopedic shoulder procedures while retaining the ability to group scoring systems based on similar outcomes, for example, PROMs that cover subluxation or shoulder dislocation can be grouped in the field of instability surgery. Our initial search included all results for “shoulder.mp” before the results were further screened. Screened articles were accessed in full-text versions and reviewed to further confirm eligibility for the study. Exclusion criteria included basic science studies, radiographic studies, cadaveric studies, pediatric studies (age less than 18 years), case reports, editorials, letters to the editor, current concept reviews, surgical techniques, and instructional course lectures. Studies were excluded when the title, abstract, or full-text version confirmed that the articles were not eligible as per the above criteria. Indicators of exclusion therefore included articles without adult patients (such as basic science articles), articles with patients but without surgical procedures, articles with patients and surgical procedures but without PROMs used to evaluate those patients, and articles detailing surgery on other areas such as the distal humerus or elbow. Data collected included the journal of publication, affiliated institutions, country of origin of the authors, the Oxford Centre for Evidence-Based Medicine (OCEBM) level of evidence,[10] the specific PROMs used, the sample size of the study at commencement and follow-up, and the statistical significance of the findings. Data were analyzed using IBM SPSS Statistics version 27.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics were performed in addition to student’s *t*-test for comparison of the PROMs usage for different journals, time periods, and regions of origin of the articles. The use of each PROM was calculated per journal and per geographical region as a function of the percentage of all articles published with that PROM. Significance was set with a P value of $< .05$, and the 95% confidence interval reported for all relevant values.

Results

We initially identified 4175 articles from the selected journals over the chosen time period. These articles were screened by title and abstract to a reduced number of 672 articles which were read in full-text format. Of the remaining 672 articles, 607 were eligible for inclusion in the study (Fig. 1). The included articles had a median OCEBM level of evidence of 3. Of the articles included, 246 (40.5%) were related to rotator cuff surgery, 242 (39.9%) were related to shoulder arthroplasty, and 119 (19.6%) were related to shoulder instability surgery. The number of eligible articles published per year is demonstrated in Figure 2, with an increase from 57 articles in 2013 to 115 articles in 2019, which corresponds to a 102% increase in the number of articles using PROMs. Over the same period of time, the total number of articles published by the combination of all selected journals rose from 2444 in 2013 to 2596 in 2019, which was a 6.2% increase in the number of journal articles published. If corrected for this 6.2% increase in journal article publications, there was only a 93% increase in the proportion of articles using PROMs over the study period. The low number of articles in 2020 reflects the cut-off point for data collection. The number of eligible articles obtained from each journal ranged from 36 (CORR) to 281 (JSES) which reflects the variation in the number of articles published by each journal and the degree to which each journal publishes specifically on shoulder surgery. The range of patient sample sizes in the articles varied from 5 to 1624, with a mean follow-up rate of 91.8% (range 13.5%-100%, standard deviation [SD] 13.36%). The range of PROMs used per article was 1-7, with a median of 3 PROMs per article (Fig. 3). Of the articles, 489 used 2 or more PROMs (80.4%). The mean number of scores used per article increased from 2.67 to 2.90 from 2013 to 2019 ($P = .24$, 95% confidence interval: −0.65 to 0.18). The total number of PROM usages recorded was 1593 which was comprised of 63 different scoring systems (Table I). The range of usages per PROM scoring system was from single use for a number of more esoteric scoring systems, to 350 usages for the American Shoulder and Elbow Surgeons Shoulder score (ASES) (Fig. 4, Table I). The top 5 most used PROMs were ASES, Constant Score, Visual Analog Scale (VAS) (pain), Simple Shoulder Test (SST), and the University of California Los Angeles (UCLA) scores (Fig. 4). These frequently used scores showed a trend for increased usage over time (Fig. 5). The percentage of all included articles using each of the top scoring systems was 57.7% of articles for ASES, 41% of articles for Constant Score, 37.1% for VAS (pain), 24.1% of articles for SST, and 16.6% of articles for the UCLA score.

The most commonly used score in articles originating from North America was the ASES score (216 uses in 273 articles; 78.1%), from Europe it was the Constant Score (129 uses in 183 articles; 70.4%), and from Asia it was the ASES score (80 uses in 126 articles; 63.4%) (Fig. 6). Only 10 of the 63 (16%) PROMs used clinician reporting as well as patient reporting in their scoring breakdown (Table I). Clinician-reported PROMs are those that require the assessment of a clinician to derive an objective value such as strength or mobility.[23] While these are not strictly ‘patient-reported’ outcome measures, they are often built into measurement tools which contain PROMs and are therefore not separated for the purposes of data presentation. The lack of clinician-reported components to the top 10 PROMs indicates that true patient reporting is preferred over combined patient and clinician reporting scales. When comparing PROMs by type of shoulder surgery, arthroplasty had 586 PROMs used in total, of which 149 (25.4%) were ASES, 95 (16.2%) were Constant Scores, and 76 (13%) were SST. Rotator cuff

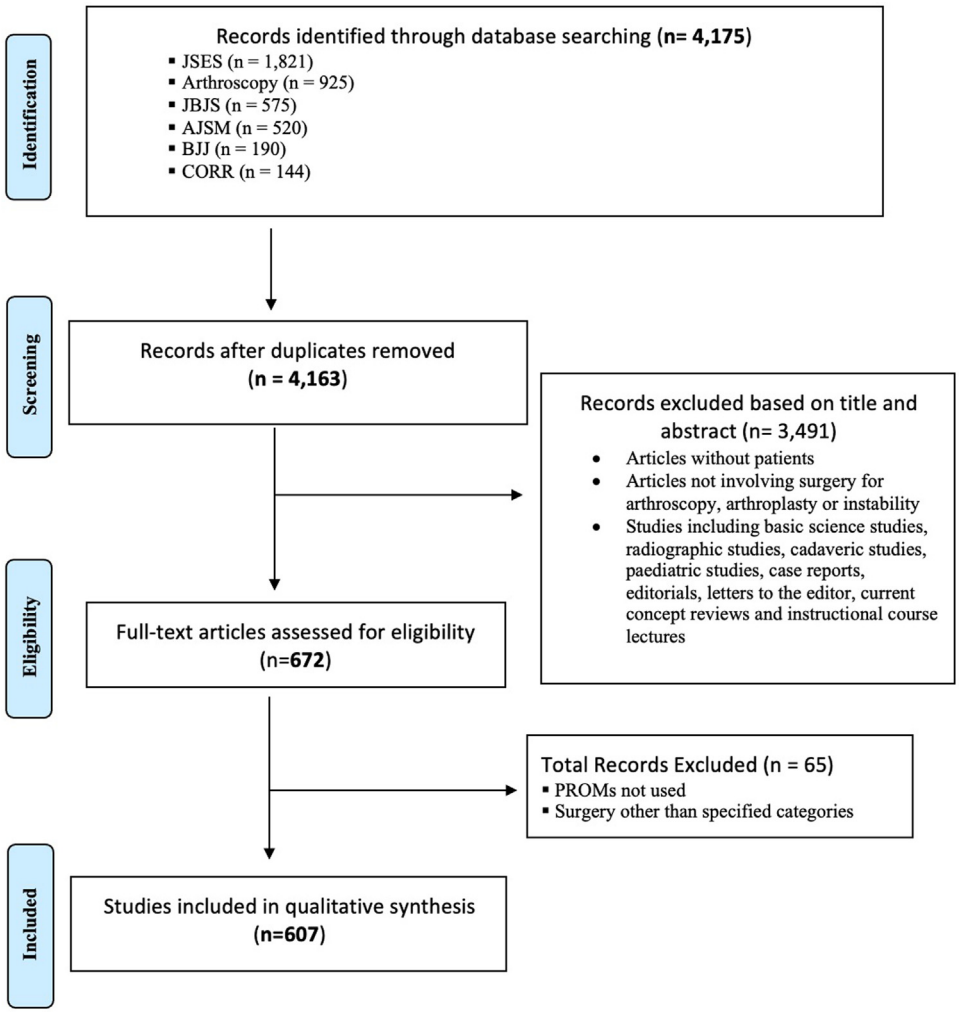

Figure 1 This PRISMA flowchart outlines the literature search process undertaken for the study. *JSES*, Journal of Shoulder and Elbow surgery; *JBJS*, Journal of Bone and Joint Surgery; *AJSM*, American Journal of Sports Medicine: *BJJ*, Bone and Joint Journal; *CORR*, Clinical Orthopaedics and Related Research; *PROMs*, patient reported outcome measures.

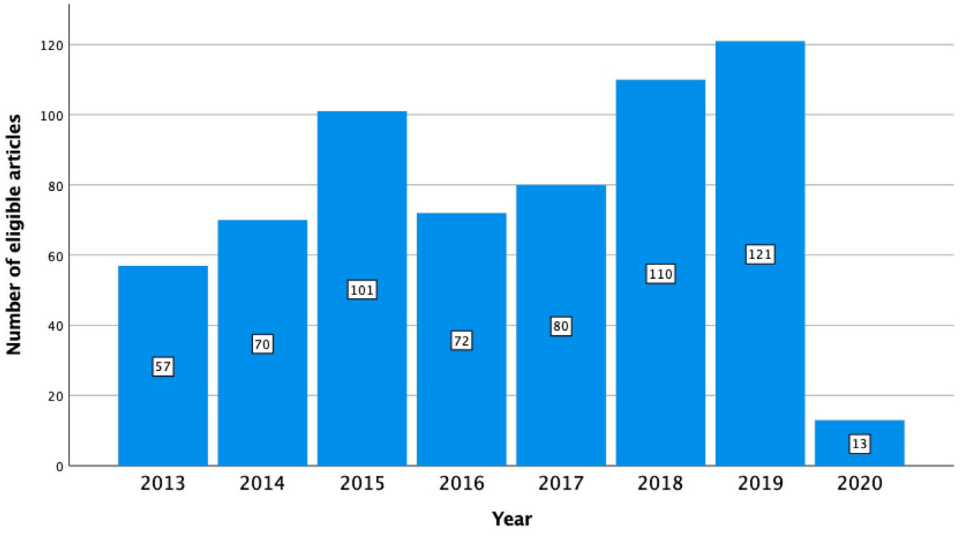

Figure 2 The blue columns in this bar graph represent the number of articles with PROMs published per year in the journals that we assessed. *PROMs*, patient reported outcome measures.

surgery had 673 PROMs used in total, of which 147 (21.8%) were ASES, 133 (19.8%) were Constant Scores, and 117 (17.4%) were VAS (pain). Instability surgery had 334 scores in total, of which ASES comprised 48 scores (14.4%), Western Ontario Shoulder instability Index (WOSI) comprised 46 scores (13.8%), and Rowe comprised 45 scores (13.5%).

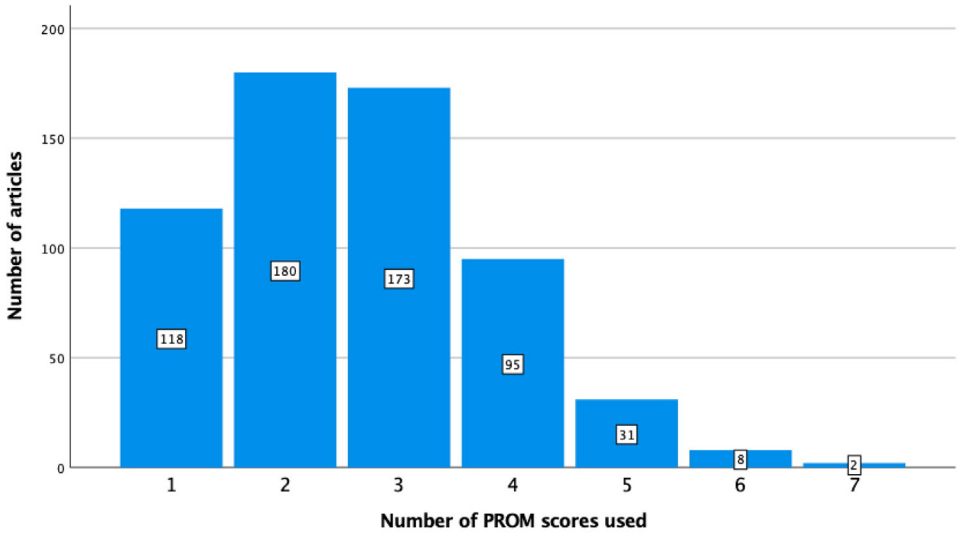

Figure 3 The blue columns in this bar graph represent the number of PROMs used per article in the articles that we assessed. *PROMs*, patient reported outcome measures.

While there was a diverse spread of PROM systems used in totality, there was strong reliance on the most commonly used PROMs. Eighty seven percent of articles used one of the top 5 most common PROMs (excluding VAS pain); 78.7% of articles which used either the ASES or Constant-Murley Score, and of these, 25.3% used both scores. This overlap was less pronounced with the fourth and fifth most commonly used PROMs; 36% of articles used either SST (fourth most commonly used PROM) or UCLA (fifth most commonly used PROM); of these, 13.3% used both scores. The mean number of PROMs per article was between 2.32 (SD 1.146) and 2.88 (SD 1.009) for all regions. There were only 2 articles that originated from Africa, 1 article from the Middle-East, and 1 article from India (Fig. 6).

Of the 63 PROMs, 36 (57.1%) reported on function only, 4 (6.3%) reported on pain only, and 23 (36.5%) reported on both measures. In the top 10 PROMs used, only 1 score assessed patient satisfaction and only 2 scores looked at psychological wellbeing (Fig. 7).

Discussion

The principal finding of this review is that between 2013 and 2019, there has been a 102% increase in the number of shoulder surgery–specific articles using PROMs, which is corrected to a 93% increase when factoring increased journal publication numbers. The number of articles using 2 or more PROMs is more than 80%, and this has increased over the time period studied. We demonstrated a high usage of well-validated, generalizable shoulder scores such as the ASES and the Constant Score, which have been shown to dominate the PROMs landscape in other reviews.[5,6,8,17] Validation is a term which infers a secondary measure to ensure that the use of a PROM is meaningful. There is no consensus on what secondary measures constitute validity, although attempts have been made to use consensus-derived standard measures.[3] We identified geographical differences in the distribution of the most commonly used PROMs in the shoulder surgery literature. The use of the Constant Score in 70.4% of captured articles published in Europe may be related to the establishment of the score as a 'gold standard' in the region, with concomitant requirement for the score to be used in meetings and publications, therefore increasing its prevalence.[20]

Comparative studies to this review have found an evolution of the use of PROMs compared to the era reported prior to the present study. A review of PROMs from Gartsman et al (2015) assessed all clinical shoulder articles published in the Journal of Bone and Joint Surgery from 2004 to 2014. Of the 575 articles published during this time, 228 were related to clinical shoulder research. Of the 228 articles, 167 (73.2%) used PROMs, with 39 outcome measurement tools used. Of the 167 articles with PROMs, there was at least 1 validated PROM used in 83.2% of articles and 2 validated PROMs used in 30.5% of articles.[6] The Constant Score was the most commonly used PROM, present in 35.9% of articles with PROMs, closely followed by ASES which was present in 31.7% of articles with PROMs. This is a lower rate of use of these top 2 PROMs than our dataset, which demonstrated the use of ASES in 57.7% of PROMs articles and the Constant Score in 41% of PROMs articles. This indicates a shift toward a more universal uptake of major PROMs compared to the Gartsman et al results.

Whittle et al (2019) reviewed PROMs in articles reporting on shoulder instability between the year 2000 and 2018. Journal articles were selected if they reported on the operative management of shoulder instability with a functional outcome assessment in patients aged 10 years and more. The review found 91 eligible articles of 4506 screened. The 91 articles contained 28 different outcome measures or scores. The Rowe score was used in 58% of articles, and the Constant Score was used in 33% of articles. The WOSI was the most commonly used score without a clinician reporting component and was used in 24% of articles. Our review demonstrated a higher rate of use of the Constant Score (41%) and a drastically lower rate of use of the Rowe score (8.1%) and WOSI score (3.1%) in PROMs articles. Part of this is due to the instability-specific nature of the Whittle et al series compared to our series which has mixed pathology. When looking at instability alone, Rowe scores account for 13.5% of PROMs used in our series. Of the 91 articles reviewed by Whittle et al,[21] the majority of PROMs used focused on pain (75%) and physical function (85%) with a much lower rate of reporting on quality of life and treatment success. Of the top 10 PROMs most used in our review, we found that the UCLA score covered the most domains, including strength, range of motion, satisfaction, and psychosocial factors (Fig. 7).

Mosher et al (2020)[15] published the largest series on upper limb PROMs to date and captured articles from 2007 to 2017. They identified 1740 articles with a total of 105 different PROM systems and an 18% increase in the use of PROMs over the study period. In comparison, our study captures less articles but captures more recent data, and our data demonstrate a much higher rate of

Table I
Patient-reported outcome measures used across all published articles.

PROM	Number of uses in publications	Percentage of journals articles Using PROM	Clinician reporting component
Activities of Daily Living Requiring External Rotation (ADLER) score	2	0.3	No
American Shoulder and Elbow Surgeons score (ASES)	350	57.7	No
Athletic Shoulder Outcome Scoring System	4	0.7	Yes
Constant	249	41	Yes
Disability of the Arm, Shoulder and Hand (DASH)	34	5.6	No
Dawson-12	1	0.2	No
Degree of Shoulder Involvement in Sports (DOSIS)	1	0.2	No
EQ-5D	5	0.8	No
Fudan University Shoulder Score	1	0.2	No
Hospital Anxiety and Depression Scale (HADS)	1	0.2	No
Japan Shoulder Society Shoulder Instability Score (JSS-SIS)	2	0.3	Yes
Japanese Orthopaedic Association (JOA)	5	0.8	Yes
Katz ADL	1	0.2	No
Kerlan-Jobe Orthopaedic Clinic Questionnaire (KJOC)	3	0.5	No
Korean Shoulder Score	5	0.8	No
L'Insalata	4	0.7	No
Long head of biceps tendon (LHB) score	1	0.2	Yes
Marx shoulder activity score	4	0.7	No
Mayo Elbow Performance Score (MEPS)	1	0.2	No
Melbourne Instability Shoulder Scale	2	0.3	No
Mental Health Inventory 5 Score (MHI-5)	1	0.2	No
MSTS (musculoskeletal tumour society)	1	0.2	Yes
Orebro musculoskeletal pain questionnaire	1	0.2	No
OSIS (Oxford Shoulder Instability Score)	9	1.5	No
Oxford Shoulder Score	34	5.6	No
Pain Catastrophizing Scale	1	0.2	No
Penn Shoulder Score	6	1.0	No
PROMIS—pain interference	1	0.2	No
PROMIS—physical function	6	1.0	No
PROMIS—upper extremity	1	0.2	No
QIDS-SR score (Quick inventory of depressive symptomatology)	1	0.2	No
Quick DASH	11	1.8	No
RAND-36	1	0.2	No
Rotator cuff quality of life index	1	0.2	No
Rowe	49	8.1	Yes
SF-12	32	5.3	No
SF-36	14	2.3	No
SF-6D	1	0.2	No
Shoulder Activity Level	1	0.2	No
Shoulder activity scale (SAS)	7	1.2	No
Shoulder Instability Score	3	0.5	Yes
Shoulder pain and disability index metrics	2	0.3	No
Shoulder Pain and Disability Index (SPADI)	19	3.1	No
Simple Shoulder Test (SST)	146	24.0	No
Single Assessment Numeric Evaluation (SANE)	59	9.7	No
Subjective Shoulder Value (SSV)	68	11.4	No
Subjective Assessment of Shoulder Function	1	0.2	No
Subjective Patient Outcome for Return to Sport (SPORT) score	1	0.2	No
Tegner activity scale	1	0.2	No
University of California Los Angeles Shoulder Score (UCLA)	101	16.6	Yes
Visual Analog Scale (VAS) (Disability)	2	0.3	No
VAS (function)	7	1.2	No
VAS (Instability)	1	0.2	No
VAS (Motion)	1	0.2	No
VAS (Pain)	225	37.1	No
VAS (QOL)	1	0.2	No
VAS (Satisfaction)	1	0.2	No
Veterans Rand 12 (VR-12)	5	0.8	No
Walch-Duplay	10	1.6	Yes
World Health Organisation (WHO) QOL BREF	1	0.2	No
WOOS (Western Ontario Osteoarthritis of the Shoulder Index)	8	1.3	No
Western Ontario Rotator Cuff Index (WORC)	19	3.1	No
Western Ontario Shoulder instability Index (WOSI)	56	9.2	No
Total	1593	-	-

PROM, patient reported outcome measures.

increase in PROMs usage (102% vs. 18%) over the study period. One reason for this difference in findings may be that the Mosher et al study looked at different journals; they used 8 journals instead of 6 and included articles from the Journal of Orthopaedic Trauma and Shoulder and Elbow (the United Kingdom). Another reason for this difference in outcomes is the methodology for the capture of PROMs; Mosher et al captured all PROMs related to shoulder conditions, including open surgical procedures, arthroscopic shoulder procedures, and nonsurgical treatment of shoulder conditions (excluding psychiatric conditions). This also explains why larger amount of PROM tools (105) was captured in the Mosher series compared to our series of 65 PROM tools. Our results also

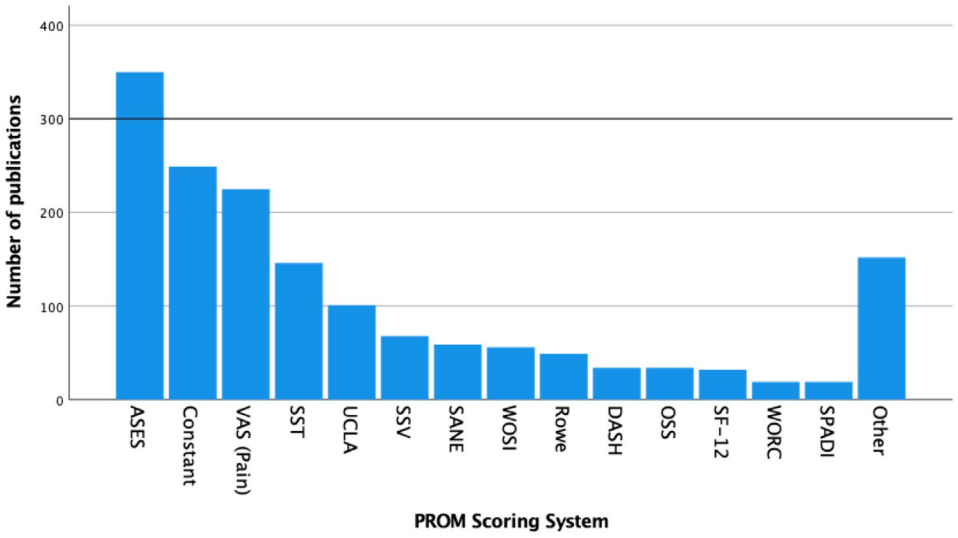

Figure 4 The blue columns in this bar graph represent the number of times each PROM system was used in a different article. *ASES*, American Shoulder and Elbow Surgeons; *VAS*, visual analog scale; *SST*, simple shoulder test; *UCLA*, University of California Los Angeles score; *SSV*, subjective shoulder value; *SANE*, single assessment numeric evaluation; *WOSI*, Western Ontario Shoulder Instability index; *DASH*, disabilities of the arm shoulder and hand score; *OSS*, Oxford Shoulder Score; *WORC*, Western Ontario Rotator Cuff index; *SPADI*, shoulder pain and disability index; *PROM*, patient reported outcome measures.

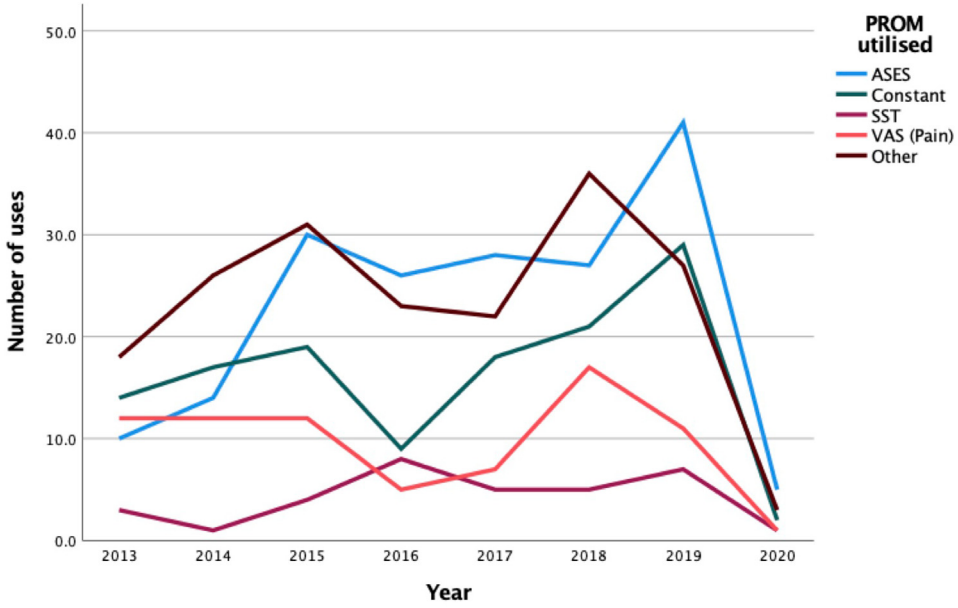

Figure 5 Each colored line in this line graph represents one of the Top 4 most commonly used PROMs and the number of times they were used in articles each year. *PROMs*, patient reported outcome measures; *ASES*, American Shoulder and Elbow Surgeons; *SST*, simple shoulder test; *VAS*, visual analog scale.

demonstrated a 44% increase in the number of PROMs used between 2017 and 2019, a period which was not captured in the Mosher et al study and also contributes to the discrepancy between the studies.

Booker et al (2015)[2] reviewed the use of PROMS in 4 major journals from 2012 to 2013 with a total of 174 articles included. They found an increasing trend toward the use of PROMs in these journals and recommended that multiple scores be used to improve accuracy. While it is important to evaluate as many of the most important effects of treatment as possible, clinicians should consider responder fatigue when performing these assessments.[7] It is perhaps prudent to identify a choice, few validated and broadly applicable PROMs to facilitate direct comparisons between similar populations of interest. As can be seen from the composition of the top 10 PROMs in use, many of the scores cover the same domains and will therefore cover duplicate areas if they are used together (Fig. 7). We found that 53.4% of articles used 2 of the 5 most common PROM systems, indicating a high potential for duplication of assessment for patient responses. In practice, this can make the use of multiple PROMs look more desirable; frequently used PROMs such as ASES and Single Assessment Numeric Evaluation have a high crossover in what they capture, which leads to high correlations between the scores when demonstrated against clinical cohorts.[22] Gowd et al (2019)[9] produced a review of the use of PROMs

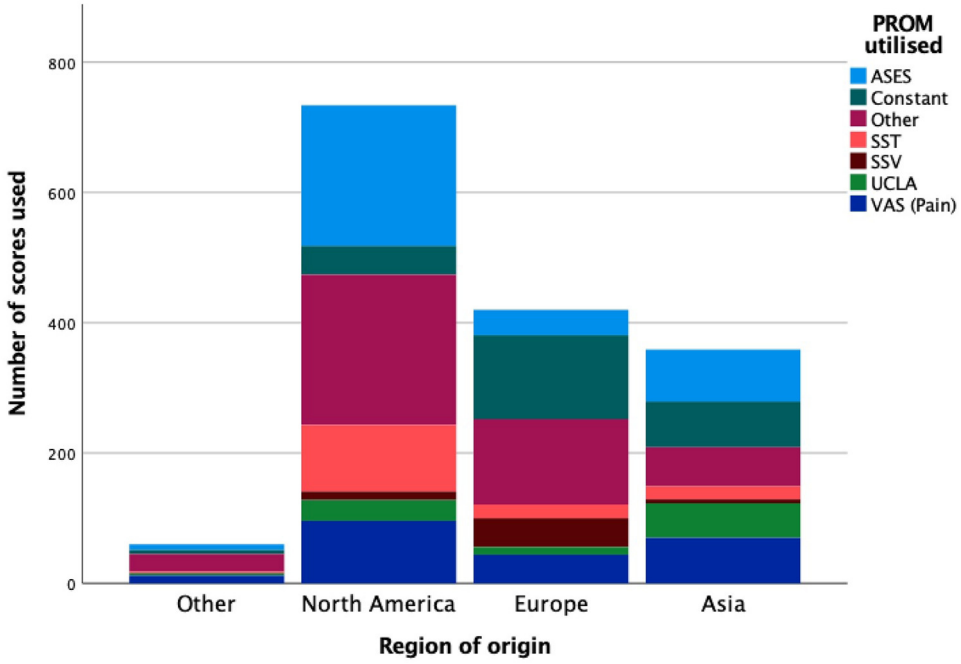

Figure 6 Each column in this bar graph represents a geographical region of origin for the published articles. The color code for each bar represents different PROMs. *PROMs*, patient reported outcome measures; *ASES*, American Shoulder and Elbow Surgeons; *SST*, simple shoulder test; *SSV*, subjective shoulder value; *UCLA*, University of California Los Angeles score; *VAS*, visual analog scale.

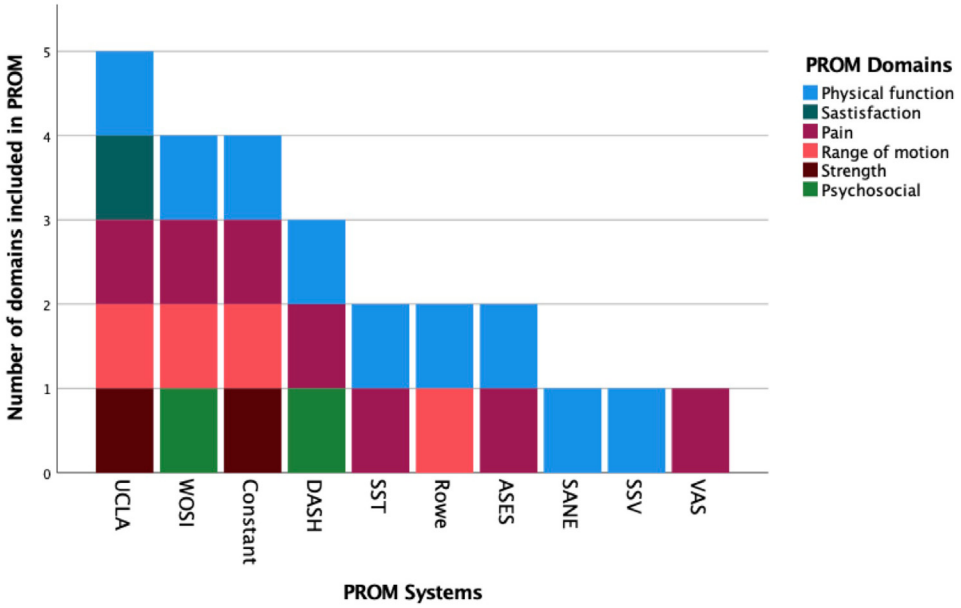

Figure 7 Each column in this bar graph represents one of the Top 10 most used PROMs. The color code for each bar represents the domains that constitute each PROM scoring system. *ROMs*, patient reported outcome measures; *UCLA*, University of California Los Angeles score; *WOSI*, Western Ontario Shoulder Instability index; *DASH*, disabilities of the arm shoulder and hand score; *SST*, simple shoulder test; *ASES*, American Shoulder and Elbow Surgeons; *SANE*, single assessment numeric evaluation; *SSV*, subjective shoulder value; *VAS*, visual analog scale.

in clavicle fracture publications, and despite reporting on a very specific injury for these patients, they suggested that to maintain the ability for cross-study comparisons, at least 2 common PROMs should be used per study, in addition to whatever more esoteric PROMs may be useful.

Despite the increasing number and variety of PROMs, they remain imperfect tools for the assessment of patient wellbeing. Jain et al (2019)[11] subcategorized PROMs into health-related quality of life scores and shoulder-related PROMs to compare their sensitivity in detecting improvements in a cohort of 145 patients undergoing anatomic total shoulder arthroplasty. They found that both sets of scores showed a significant improvement in patient function from baseline, although the health-related quality of life scores such as the EQ-5D, SF-6D, and VAS quality of life had only weak correlations with the outcomes of the shoulder-specific PROMs such as the ASES and DASH. An ideal PROM has limited ceiling and floor effects,

meaning that it can measure the extremes of poor or excellent function without reaching saturation in its scoring and sustaining a loss of distinction between patients clustered with similar outlier scores.[12] The more commonly used PROMs identified in this article have all been validated against multiple clinical cohorts, this includes ASES, Single Assessment Numeric Evaluation, UCLA, Oxford Shoulder Score, Constant, Rowe, SST, Shoulder Pain and Disability Index, Penn Shoulder Score, and WORC.[4,5,14,16,18,19] Even well-validated PROMs such as the ASES has the issue of catering to the functionality of high-demand patients (with questions regarding overhead throwing), which is not a relevant measure of function for most patients aged more than 65 years.[5] The utility of our results in the current context of shoulder surgery is in demonstrating the expansion of the field of PROMs and highlighting the prevalent PROMs used per region. On first review, it may appear that there is an excessive number of PROMs being used in the shoulder literature (Table I); however, 53.4% of the articles use at least 2 PROMs from the top 5, and 78% of the articles use the ASES score or the Constant Score and 25.3% use both, which allows a good comparison of the results between the series. The results of this study should be used to streamline the PROMs selection process for researchers wishing to measure PROMs for upper limb surgery. This should be independent of geographical stratification; shoulder surgery patient reporting should cross geographical, language, and cultural boundaries to allow a global synthesis of efficacy of surgery. We recommend that the efficacy of surgery should be measured with a clinical and a quality of life component and with PROMs that are stratified to target patients with specific shoulder pathology.

The main strength of this review was in its capture of a large number of articles (n = 607) over a 7-year period. The mean impact factor of the journals included in this study was 4.36, which denotes a high level of confidence in the peer-review of the articles that were included. The included articles were predominately studies with a high level of evidence; the studies had a median OCEBM level of evidence of 3, with 52.8% of the articles being level 3 or more. The limitations of this review are predominately of its finite data capture period and scope. This review period captures PROMs data in the pre-COVID era but not the post-COVID period. This review also only captures data from major journals, which creates a selection bias for the findings. We acknowledge that the findings of this study are not necessarily applicable to all recent published shoulder research, but the findings are relevant at the top-end of the literature where there is maximal peer review and a high number of citations, and therefore a strong influence on the orthopedic community. While subdividing shoulder surgery into 3 fields was useful for the purpose of analysis, this necessarily fails to capture procedures that fall outside of these categories, such as open tumor resection or open reduction and internal fixation of proximal humerus fractures. Our capture of PROMs use is limited to journal, time, and location; future studies may be able to use more sophisticated methodology to assess trends in PROM usage.

Conclusion

Our study shows there is a boom in the use of PROMs for upper limb surgery research, and with this carries the risk of over-expansion of the field of PROMs. Our concern with this is the possibility of creation of more data that lack direct comparability to that of the rest of the scientific community. Published reviews of PROMs in the literature are commensurate with our findings in terms of the increasing prevalence of, and diversity of PROMs systems being used in upper limb surgery. We found that there was a paucity of clinician reporting for the PROMs systems used, with a much higher proportion of PROMs used that were patient-reported only. There is a continental preference for the ASES score in North America and Asia, and a preference for the Constant Score in Europe, which may be related to historical requirements for the use of this score in European institutions. Despite there being a clear preference for the ASES, Constant, and VAS scores, these measures are relatively blunt instruments which do not take into account patient satisfaction or psychosocial factors.

For future studies we cannot recommend that the most commonly used PROMs be reproduced for any given shoulder surgery study. Each clinical study involving shoulder surgery will need to answer specific questions, and it is advisable to use the full scope of the PROMs resources to find the best PROM to sensitively detect the differences in outcomes that are hypothesized. Once this condition has been met, if there is further choice, it is preferable to use PROMs which are validated and can be used to add directly comparable results to the canon of upper limb surgery literature.

Acknowledgment

The authors would like to acknowledge Dr. Laurant Kang, MBBS who contributed to the data collection for this article but his contribution did not justify authorship.

Disclaimers:

Funding: The corresponding author is supported by an Australian Government Research Training Program Scholarship. No direct funding was received for this study in the form of grants, equipment, or other items for this study.

Conflicts of interest: Dr. Rajpal Narulla is supported by an Australian Government Research Training Program Scholarship. The other authors, their immediate families, and any research foundation with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

References

1. An VVG, Farey JE, Karunaratne S, Smithers CJ, Petchell JF. Subacromial analgesia via continuous infusion catheter vs. placebo following arthroscopic shoulder surgery: a systematic review and meta-analysis of randomized trials. J Shoulder Elbow Surg 2020;29:471-82. https://doi.org/10.1016/j.jse.2019.11.010.
2. Booker S. Use of scoring systems for assessing and reporting the outcome results from shoulder surgery and arthroplasty. World J Orthop 2015;6:244. https://doi.org/10.5312/wjo.v6.i2.244.
3. Churruca K, Pomare C, Ellis LA, Long JC, Henderson SB, Murphy LED, et al. Patient-reported outcome measures (PROMs): a review of generic and condition-specific measures and a discussion of trends and issues. Health Expect 2021;24:1015-24. https://doi.org/10.1111/hex.13254.
4. Cvetanovich GL, Gowd AK, Liu JN, Nwachukwu BU, Cabarcas BC, Cole BJ, et al. Establishing clinically significant outcome after arthroscopic rotator cuff repair. J Shoulder Elbow Surg 2019;28:939-48. https://doi.org/10.1016/j.jse.2018.10.013.
5. Fu MC, Chang B, Wong AC, Nwachukwu BU, Warren RF, Dines DM, et al. PROMIS physical function underperforms psychometrically relative to American Shoulder and Elbow Surgeons score in patients undergoing anatomic total shoulder arthroplasty. J Shoulder Elbow Surg 2019;28:1809-15. https://doi.org/10.1016/j.jse.2019.02.011.
6. Gartsman GM, Morris BJ, Unger RZ, Laughlin MS, Elkousy HA, Edwards TB. Characteristics of clinical shoulder research over the last decade: a review of shoulder articles in The Journal of Bone & Joint Surgery from 2004 to 2014. J Bone Joint Surg 2015;97:e26. https://doi.org/10.2106/JBJS.N.00831.
7. Gibbons C, Porter I, Gonçalves-Bradley DC, Stoilov S, Ricci-Cabello I, Tsangaris E, et al. Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. Cochrane Database Syst Rev 2021;10:CD011589. https://doi.org/10.1002/14651858.CD011589.pub2.
8. Gowd AK, Charles MD, Liu JN, Lalehzarian SP, Cabarcas BC, Manderle BJ, et al. Single Assessment Numeric Evaluation (SANE) is a reliable metric to measure

clinically significant improvements following shoulder arthroplasty. J Shoulder Elbow Surg 2019;28:2238-46. https://doi.org/10.1016/j.jse.2019.04.041.
9. Hao KA, Kakalecik J, Delgado GA, Wright TW, King JJ, Wright JO. Current trends in patient-reported outcome measures for clavicle fractures: a focused systematic review of eleven influential orthopaedic journals. J Shoulder Elbow Surg 2021;31:e58-67. https://doi.org/10.1016/j.jse.2021.08.034.
10. Howick J, Chalmers I, Glasziou P, Greenhaigh T, Heneghan C, Liberati A, et al. Explanation of the 2011 Oxford Centre for evidence-based Medicine (OCEBM) levels of evidence (background Document). Oxford Centre for Evidence-Based Medicine. https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence. Accessed March 22, 2022.
11. Jain SS, DeFroda SF, Paxton ES, Green A. Patient-reported outcome measures and health-related quality-of-life scores of patients undergoing anatomic total shoulder arthroplasty. J Bone Joint Surg 2020;102:1186-93. https://doi.org/10.2106/JBJS.20.00402.
12. 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:e021. https://doi.org/10.5435/JAAOS-Global-D-17-00021.
13. Karunaratne S, Duan M, Pappas E, Fritsch B, Boyle R, Gupta S, et al. The effectiveness of robotic hip and knee arthroplasty on patient-reported outcomes: a systematic review and meta-analysis. Int Orthop 2019;43:1283-95. https://doi.org/10.1007/s00264-018-4140-3.
14. Michael RJ, Williams BA, Laguerre MD, Struk AM, Schoch BS, Wright TW, et al. Correlation of multiple patient-reported outcome measures across follow-up in patients undergoing primary shoulder arthroplasty. J Shoulder Elbow Surg 2019;28:1869-76. https://doi.org/10.1016/j.jse.2019.02.023.
15. Mosher ZA, Ewing MA, Collins CS, Young PG, Brabston EW, Momaya AM, et al. Usage trends of patient-reported outcome measures in shoulder literature. J Am Acad Orthop Surg 2020;28:e774-81. https://doi.org/10.5435/JAAOS-D-19-00455.
16. Park I, Oh M-J, Shin S-J. Minimal clinically important differences and correlating factors for the Rowe score and the American Shoulder and Elbow Surgeons score after arthroscopic stabilization surgery for anterior shoulder instability. Arthroscopy 2019;35:54-9. https://doi.org/10.1016/j.arthro.2018.08.005.
17. Retzky JS, Baker M, Hannan CV, Srikumaran U. Single assessment numeric evaluation scores correlate positively with American Shoulder and Elbow Surgeons scores postoperatively in patients undergoing rotator cuff repair. J Shoulder Elbow Surg 2020;29:146-9. https://doi.org/10.1016/j.jse.2019.05.039.
18. Sahoo S, Ricchetti ET, Zajichek A, , Cleveland Clinic Shoulder Group, Evans PJ, Farrow LD, et al. Associations of preoperative patient mental health and sociodemographic and clinical characteristics with baseline pain, function, and satisfaction in patients undergoing rotator cuff repairs. Am J Sports Med 2020;48:432-43. https://doi.org/10.1177/0363546519892570.
19. Strong B, Maloney M, Baumhauer J, Schaffer J, Houck JR, Hung M, et al. Psychometric evaluation of the patient-reported outcomes measurement Information system (PROMIS) physical function and pain interference computer adaptive test for subacromial impingement syndrome. J Shoulder Elbow Surg 2019;28:324-9. https://doi.org/10.1016/j.jse.2018.07.024.
20. Tuton D, Barbe C, Salmon J-H, Dramé M, Nérot C, Ohl X. Transcultural validation of the oxford shoulder score for the French-speaking population. Orthop Traumatol Surg Res 2016;102:555-8. https://doi.org/10.1016/j.otsr.2016.05.010.
21. Whittle JH, Peters SE, Manzanero S, Duke PF. A systematic review of patient-reported outcome measures used in shoulder instability research. J Shoulder Elbow Surg 2020;29:381-91. https://doi.org/10.1016/j.jse.2019.07.001.
22. Wickman JR, Lau BC, Scribani MB, Wittstein JR. Single Assessment Numeric evaluation (SANE) correlates with American shoulder and Elbow Surgeons score and Western Ontario rotator cuff index in patients undergoing arthroscopic rotator cuff repair. J Shoulder Elbow Surg 2020;29:363-9. https://doi.org/10.1016/j.jse.2019.07.013.
23. Wolfensberger A, Vuistiner P, Konzelmann M, Plomb-Holmes C, Léger B, Luthi F. Clinician and patient-reported outcomes are associated with psychological factors in patients with chronic shoulder pain. Clin Orthop Relat Res 2016;474:2030-9. https://doi.org/10.1007/s11999-016-4894-0.
24. Xu S, Chen JY, Lie HME, Hao Y, Lie DTT. Determination of threshold scores for treatment success after arthroscopic rotator cuff repair using Oxford, constant, and University of California, Los Angeles Shoulder scores. Arthroscopy 2019;35:304-11. https://doi.org/10.1016/j.arthro.2018.07.047.