J Shoulder Elbow Surg (2021) 30, 1174–1180

ELSEVIER

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

Responsiveness of patient-reported outcomes in shoulder arthroplasty: what are we actually measuring?

Kevin J. Cronin, MD, MS[a,*], Justin A. Magnuson, BA[a], Meredith L. Murphy, BS[a], R. Zackary Unger, MD[a], Cale A. Jacobs, PhD[a], Matthew H. Blake, MD[b]

[a]*Department of Orthopaedic Surgery & Sports Medicine, University of Kentucky, Lexington, KY, USA*
[b]*Avera Orthopedics, Sioux Falls, SD, USA*

Hypothesis: The purpose of this study was to determine the prevalence and responsiveness of common patient-reported outcome (PRO) tools in patients undergoing primary total shoulder arthroplasty (TSA) for glenohumeral arthritis.
Methods: Using Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines, a systematic review of anatomic and reverse TSA studies from PubMed, SportDiscus, Cochrane, and CINAHL was performed. Studies on primary TSA for glenohumeral arthritis that reported at least 1 PRO tool were included in the final analysis. A subgroup analysis of studies that reported preoperative and postoperative PRO scores with at least 2-year follow-up data was evaluated to compare the responsiveness between the different PRO instruments.
Results: After full-text review of 490 articles, 74 articles met all inclusion criteria and were included in the final analysis. Anatomic TSA was evaluated in 35 studies, reverse TSA in 32 studies, and both anatomic and reverse in 7 studies. There were a total of 7624 patients, and 25 different PRO tools were used. The most commonly reported PRO tools were the American Shoulder and Elbow Surgeons (44 studies), Constant (42 studies), the visual analog scale for pain (23 studies), and the Simple Shoulder Test (17 studies). A median of 3.0 PRO instruments were used in each study. All instruments had large effect sizes. The University of California at Los Angeles (UCLA) score was found to be the most responsive instrument, and the Single Assessment Numeric Evaluation score was least responsive. The American Shoulder and Elbow Surgeons score was the most responsive instrument that required only patient-reported data.
Conclusion: Overall, the UCLA score was found to be the most responsive followed by the Adjusted Constant. However, both the UCLA and Adjusted Constant scores require strength and range of motion assessment that may limit their widespread clinical use. The increased responsiveness of these measures, which include objective clinical testing, speaks to the predicted increases in strength and range of motion after shoulder arthroplasty. Of the measures that can be administered without in-person clinical evaluation, the American Shoulder and Elbow Surgeons score and Western Ontario Osteoarthritis of the Shoulder index were the most responsive.
Level of evidence: Meta-analysis; Basic Science Study; Validation of Outcome Instruments

Keywords: Shoulder; shoulder arthroplasty; patient-reported outcomes; PROs; reverse total shoulder arthroplasty; anatomic total shoulder arthroplasty

Institutional review board approval was not required for this basic science study.

*Reprint requests: Kevin J. Cronin, MD, MS, Department of Orthopaedic Surgery & Sports Medicine, University of Kentucky, 740 S. Limestone, K403, Lexington, KY 40536, USA.
E-mail address: kevincronin88@gmail.com (K.J. Cronin).

Patient-reported outcomes (PROs) have become increasingly critical components of evaluating a wide variety of orthopedic conditions and procedures, having utility in both research and clinical practice.[8,23] In addition, a growing emphasis on patient satisfaction and cost-

1058-2746/$ - see front matter
https://doi.org/10.1016/j.jse.2020.08.019

effectiveness within the health care system leads to an even greater need to accurately and efficiently evaluate patient outcomes.[2,21] A multitude of validated PROs exist for the shoulder, ranging from general shoulder assessments to disease-specific tools, including measures specific for instability, rotator cuff pathology, and glenohumeral arthritis.[15,28] Selecting the most appropriate instrument to evaluate a specific patient population remains a challenge for clinicians and researchers.[9]

Responsiveness is a measure of a PRO tool's ability to accurately and efficiently reflect a change within a population over time and is an important psychometric property.[3,13] Measuring responsiveness allows comparison of multiple tools to determine the optimal combination of PROs to evaluate outcomes of specific interest.[6] The relative responsiveness of PROs for rotator cuff and shoulder instability has been described in the literature; however, responsiveness for PROs used to evaluate patients undergoing shoulder arthroplasty is limited.[27] The purpose of this study was to determine the prevalence and responsiveness of common PROs used to evaluate patients undergoing primary anatomic or reverse total shoulder arthroplasty (TSA) for glenohumeral arthritis based on a systematic review of the literature.

Methods

Systematic review

Using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines, a systematic review of anatomic and reverse TSA studies was conducted. The PubMed, SportDiscus, Cochrane, and CINAHL databases were searched using the terms "shoulder arthroplasty" and "outcomes" and the terms "shoulder arthroplasty" and "patient reported outcomes."[19] Included publications were limited to English language studies on human subjects assessing outcomes after primary TSA for glenohumeral arthritis published between May 18, 2007, and May 17, 2017. Exclusion criteria included studies that did not use standardized forms of PROs, studies that were not primary research (such as review articles or case reports), anatomic/technique articles, studies on arthroplasty for the treatment of fractures or revision arthroplasty, studies without a full-text manuscript, and articles not available in English.

Subgroup statistical analysis

For each eligible study, the PROs used, number of patients, mean follow-up time, and the preoperative and postoperative means and standard deviations were recorded. A subgroup analysis of studies that reported preoperative and postoperative PRO scores with at least 2-year follow-up data was evaluated to compare the responsiveness between the different PRO instruments. Only articles using at least 2 PROs and reporting preoperative and postoperative means and standard deviations were able to be used for this subgroup analysis.

Effect size and relative efficiency (RE) were calculated and used to directly compare the responsiveness between PROs.[16,20,27] Effect size, determined individually for each PRO, was calculated by dividing the difference between postoperative and preoperative mean scores by the preoperative standard deviation. Effect size is a measure of magnitude of change within the tool after surgical intervention. An effect size is considered small if it is between 0.20 and 0.49, moderate if between 0.50 and 0.79, and large if $\geq$0.80.[7]

RE, determined from articles reporting multiple PROs, was calculated to compare each tool. RE was calculated by dividing the t score from paired t tests of one PRO by the t score of other PROs within the same study and squaring the results. An RE of >1 for one PRO compared with another indicates that the initial tool is "more responsive" than second.[17,20] Inversely, an RE of <1 indicates that a PRO is "less responsive" than a second to which it is being compared.

Results

Prevalence of PROs

Initial search yielded 687 abstracts. Duplicates were then removed, leaving 490 unique articles. After screening by title and abstract, 144 articles were identified for full-text review, of which 74 met inclusion criteria (Fig. 1). Thirty-five articles evaluated anatomic TSA, 32 evaluated reverse TSA, and 7 evaluated both, with a total of 7694 procedures included in all studies. Twenty-five unique PRO tools were reported, with a mean of 3.0 per study. The American Shoulder and Elbow Surgeons (ASES) score was the most common (44 studies), followed by the Constant score (42 studies), the visual analog scale for pain (23 studies), and the Simple Shoulder Test (SST, 17 studies) (Table I).

Subgroup analysis

Sixteen articles reported pre- and postoperative PRO means and standard deviations and were included in the subgroup analysis (Supplementary Appendix S1). These 16 studies included a total of 2488 patients. A total of 10 PROs were reported with a mean of 4.0 PROs per study (Table II). The ASES and Constant were again the most common instruments (12 studies each), followed by the Adjusted Constant (5 studies), Western Ontario Osteoarthritis of the Shoulder index (WOOS) (5 studies), SST (4 studies), University of California at Los Angeles score (UCLA) (3 studies), Shoulder Pain and Disability Index (3 studies), Single Assessment Numeric Evaluation (SANE) (3 studies), and Subjective Shoulder Value (SSV) (2 studies). The Oxford Shoulder Score and Activities of Daily Living that require active External Rotation score were used by a single study each within the subgroup, and were thus excluded from responsiveness analysis.

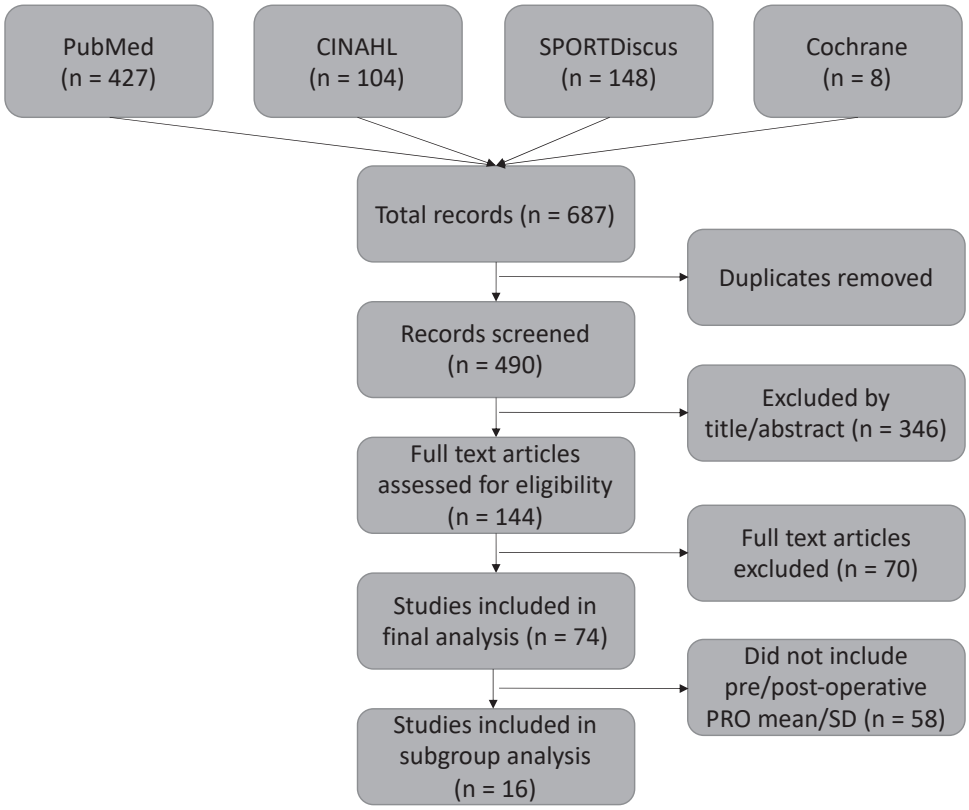

Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of article inclusion. *PRO*, patient-reported outcome; *SD*, standard deviation.

Table I Prevalence of PROs in all 74 studies identified

PRO	Number of studies	PRO	Number of studies
ASES	44	DASH	3
Constant	42	VAS function	3
VAS pain	23	Penn	2
SST	17	SF-36 MCS	2
Adjusted Constant	14	SF-36 PCS	2
SPADI	8	ADLER	1
UCLA	8	DSST	1
WOOS	8	EQ5D	1
SSV	7	MACTAR	1
OSS	6	QuickDASH	1
SF-12 PCS	6	Penn Function	1
SANE	5	Penn Pain	1
SF-12 MCS	5		

PRO, patient-reported outcome; *ASES*, American Shoulder and Elbow Surgeons score; *VAS*, visual analog score; *SST*, Simple Shoulder Test; *SPADI*, Shoulder Pain and Disability Index; *UCLA*, University of California at Los Angeles score; *WOOS*, Western Ontario Osteoarthritis of the Shoulder index; *SSV*, Subjective Shoulder Value; *OSS*, Oxford Shoulder Score; *SF-12*, 12-Item Short Form Health Survey; *PCS*, Physical Component Summary; *SANE*, Single Assessment Numeric Evaluation; *MCS*, Mental Component Summary; *DASH*, Disabilities of the Arm, Shoulder, and Hand score; *Penn*, Penn Shoulder Score; *SF-36*, 36-Item Short Form Survey; *ADLER*, Activities of Daily Living that require active External Rotation; *DSST*, Dutch Simple Shoulder Test; *EQ-5D*, EuroQoL 5 Dimensions; *MACTAR*, McMaster Toronto Arthritis patient preference questionnaire.

Table II PROs evaluated in subgroup analysis

PRO	Number of studies	Total number of patients	Preoperative mean	Postoperative mean	Effect size
ASES	12	2222	33.5	78.2	2.7
Constant	12	1898	32.4	70.6	2.6
Adjusted Constant	5	355	38.5	94.8	2.6
WOOS	5	248	25.7	76.8	3.2
SST	4	1463	3.3	10.0	2.4
SANE	3	175	28.6	61.3	2.8
SPADI	3	1278	18.5	79.3	2.8
UCLA	3	1256	13.0	30.0	4.0
SSV	2	223	24.1	75.1	3.2

PRO, patient-reported outcome; *ASES*, American Shoulder and Elbow Surgeons score; *WOOS*, Western Ontario Osteoarthritis of the Shoulder index; *SST*, Simple Shoulder Test; *SANE*, Single Assessment Numeric Evaluation; *SPADI*, Shoulder Pain and Disability Index; *UCLA*, University of California at Los Angeles score; *SSV*, Subjective Shoulder Value.
Higher means represent better scores in all PROs—those with multiple scales (WOOS, SPADI) converted to this format.

Responsiveness analysis

The 9 PROs used in multiple studies were included in the subgroup analysis of responsiveness. All PROs showed effect sizes >0.8, demonstrating the ability of each to reflect a large magnitude of change after surgery (Table II). Based on RE, the UCLA and Adjusted Constant scores were the most responsive instruments, with positive RE values against each PRO to which they were compared (Table III). The SSV had positive RE values compared with the ASES and Constant scores, though its limited use in the literature only allowed for comparison against these 2 other instruments. The SANE had lower RE values than every other tool with which it was compared. Between the 2 most commonly used PROs, the ASES score had a higher RE than the Constant score. The SST, another of the most commonly used PROs, had a lower RE than every PRO with which it was compared except for the Constant. Of the PROs that do not require physician-inputted data (ASES, WOOS, SSV, SANE, SST, and Shoulder Pain and Disability Index), all but the WOOS and SANE were more responsive than the Constant. Overall, the Constant was the lowest performing physician-dependent PRO when compared with entirely patient-reported PROs (Table III).

Discussion

Our review found that the most common PROs for evaluating outcomes after anatomic and reverse TSA for glenohumeral arthritis include the ASES, Constant, and SST scores. Although studies generally used multiple

Table III Relative efficiency of individual comparisons of PROs

	ASES	Constant	Adjusted Constant	SST	WOOS	SPADI	UCLA	SANE	SSV
ASES		1.15	0.90	1.28	0.75	1.08	0.65	2.96	0.62
Constant	0.87		0.84	0.91	1.82	0.92	0.57	6.37	0.79
Adjusted Constant	1.12	1.19		1.15	1.89	–	–	6.63	–
SST	0.78	1.09	0.87		–	0.96	0.57	–	–
WOOS	1.33	0.55	0.53	–		–	–	3.51	–
SPADI	0.92	1.09	–	1.05	–		0.60	–	–
UCLA	1.54	1.75	–	1.75	–	1.67		–	–
SANE	0.34	0.16	0.15	–	0.29	–	–		–
SSV	1.60	1.27	–	–	–	–	–	–	

PRO, patient-reported outcome; *ASES*, American Shoulder and Elbow Surgeons score; *SST*, Simple Shoulder Test; *WOOS*, Western Ontario Osteoarthritis of the Shoulder index; *SPADI*, Shoulder Pain and Disability Index; *UCLA*, University of California at Los Angeles score; *SANE*, Single Assessment Numeric Evaluation; *SSV*, Subjective Shoulder Value.
Values >1 indicate that PRO in the left column has greater relative efficiency compared with PRO in the top row; values <1 indicate that PRO in the left column has lesser relative efficiency than PRO in the top row; "–" indicates that PROs were unable to be compared as they were not used together in any studies analyzed.

instruments, there is currently no consensus on the most appropriate tools. Only the ASES (56%) and Constant (53%) scores were reported by more than half of the 74 articles included. Although all PROs demonstrated a large effect size, the UCLA and Adjusted Constant scores showed the highest RE compared with other instruments, suggesting that they are among the most responsive tools. Of the 2 most commonly used PROs, the ASES was more responsive than Constant.

Multiple factors contribute to the decision of which PRO to use when evaluating a treatment, including the underlying pathology, ease of administration, and surgeon preference. In a study using similar methodology to compare PROs for rotator cuff tears and shoulder instability, it was found that responsiveness is dependent on the specific condition being evaluated.[27] Another significant factor in collecting outcomes data is the time burden on both providers and patients, which varies widely between instruments. The disease-specific WOOS, for example, consists of 19 items evaluating pain and function, whereas the SANE and SSV each consist of a single item evaluating the patients' current perception of their shoulder as a percentage of a totally normal shoulder.[10,15,28] The Constant and UCLA scores consist of 8 and 5 items, respectively, but require clinician-based measurements of strength and range of motion, necessitating an office visit to obtain a score.[1,5]

The ASES, one of the most commonly used instruments, was originally described with a clinician-completed physical examination component but is often scored using only the patient-reported portion.[22,28] Advantages to PROs that are completely patient derived—including the ASES along with the aforementioned WOOS, SANE, and SSV—include simplicity of administration and less risk of inter- and intraobserver biases. Another advantage of entirely patient-reported information is the ability to collect information remotely without an in-person visit with a provider. Subsequently, there is a greater ability to capture data at more timepoints with less cost to patients and providers, while still attaining high rates of follow-up.[4]

Shoulder arthritis can be a debilitating disease and goals of arthroplasty include improving function and pain. However, controversy remains on how to best assess and measure these goals. Matsen et al[18] recently showed correlation between range of motion measurements and the SST, which is completely patient derived, to be highly variable in patients with arthritis, concluding that PROs and clinical measurements should be treated as 2 important yet distinct entities. In addition, they noted a large variation in the range of motion required to perform shoulder functions assessed by PROs, suggesting that this is often individual specific and may differ by as much as 30°-40° between patients for the same function.[18] Goodman et al[11] found similar results when comparing the patient-reported component and the physician-reported component of the ASES in patients undergoing reverse TSA, noting that only improvement in active forward flexion correlated with ASES improvement at 2 years. These discrepancies indicate that PROs attempting to quantify pain and functional deficits of the shoulder should ideally consist of entirely patient-reported data, with clinical measurements assessed and interpreted separately.

The 2 highest performing instruments by RE identified in our study, the UCLA and Adjusted Constant scores, both require clinician-reported data. The UCLA score allocates points to both patient-reported pain and function assessment along with physician-reported strength and range of motion testing, although places more emphasis on the patient-reported component.[15] The UCLA is also a short assessment that can be quickly completed by both the patient and physician. Although widely used due to being one of the first described shoulder outcome measures, it has yet to be formally validated, and no minimal clinically important difference has been established, further complicating interpretation of values. The Adjusted Constant, determined by the original Constant tool normalized to age- and sex-matched controls, is based on the rationale that range of motion may differ by gender and naturally deteriorate with age.[14] The results of our study suggest that this tool is more responsive than raw scores reported using the original Constant. The considerably lower number of overall patients evaluated with the Adjusted Constant indicate that the original Constant is much more widely used, but there may be the increased interpretive value in adjusted scores.

Another highly performing PRO, albeit with a relatively low number of patients, was the SSV. The SSV, a single question evaluation (the current state of the patients' shoulder expressed as a percentage of normal), showed high RE, although could only be compared with the ASES and Constant scores due to a limited number of studies using the score. The SANE, which is also a single question asking the patients to rate the state of their shoulder from 0 to 100, showed a lower RE compared with the ASES, Constant, Adjusted Constant, and WOOS. The strikingly different performances of the SSV and SANE, despite their similar structure along with the low overall number of patients evaluated, suggest that more direct comparisons are necessary to more fully characterize their responsiveness. Thigpen et al[26] showed that the SANE performed well compared with the ASES, with similar reliability and validity to the ASES. The low burden of completion of the SANE and SSV is a primary driver of their value, particularly in a setting such as clinical evaluation without the goal of robust data collection for research purposes.

Sciascia et al[24] performed a similar analysis comparing the responsiveness of PROs within a single cohort of patients undergoing anatomic TSA for arthritis. They also found large effect sizes for 4 instruments evaluated, noting the Constant score to have the highest RE, followed by the

WOOS, ASES, and SANE. Our pooled analysis similarly showed the SANE to be among the least responsive. However, we found the ASES to have a higher RE than the Constant but less so than the WOOS. This is the only previous study we are aware of directly comparing responsiveness among shoulder arthroplasty patients.

Our study highlights that high prevalence of use does not necessarily equal high responsiveness. Primarily, the Constant was one of the 2 most commonly used PROs, but was found to have lower responsiveness than every PRO other than the WOOS and SANE, similar to findings in patients with rotator cuff pathology and shoulder instability.[27] The ASES Value Committee recently suggested that PROs for clinical assessment of the shoulder should include the Veterans Rand 12 as a measure of general health, the ASES score, and the SANE, with additional disease-specific measures included for research purposes when applicable.[12] Their recommendations focused exclusively on PROs derived entirely from patient-reported data. Our study provides further details regarding the responsiveness and value of instruments specific to the setting of shoulder arthroplasty.

Disease-specific instruments, such as the WOOS, may add valuable additional data on top of a standardized set of PROs for all patients with shoulder-related complaints. The WOOS—a tool specific for glenohumeral osteoarthritis—performed worse than the Constant and Adjusted Constant scores, although had a higher RE than the ASES, another purely patient-reported but general evaluation of the shoulder. The WOOS is longer than ASES, and therefore would likely be included in attempts to gather more extensive data for research purposes. Our study found that the WOOS was not one of the most commonly used PROs, again demonstrating that prevalence and responsiveness are independent entities.

Limitations general to any systematic review also apply to this study. The quality of individual studies included in this review limits the quality of the review itself. Quality assessments were not required in the inclusion criteria in an attempt to broaden search efforts, and thus, any biases in the analyzed studies may underlie findings from this study. Limitations specific to our study derive from variation in the number of patients evaluated with each tool and the combinations of instruments selected by individual studies. Using this method of calculating RE, only studies using multiple PROs were eligible for analysis, thus resulting in the inability to calculate certain combinations. The interpretation of RE may also be viewed as subjective, as there is no clear threshold at which RE values represent a significant difference. A similar study of hip arthroscopy PROs used values of <0.80, 0.80-1.20, and >1.20 to define a lower RE, equivocal RE, and greater RE, respectively.[25] Some PROs were used to evaluate significantly fewer patients, limiting our ability to draw definitive conclusions. In addition, only articles reporting pre- and postoperative means and standard deviations could be included in analysis, resulting in a lower number of articles contributing to our pooled calculations.

Conclusion

Many PRO options exist for the evaluation of patients undergoing TSA for glenohumeral arthritis. No clear best PRO exists, although responsiveness should be considered when selecting instruments for tracking clinical outcomes, along with factors such as ease of administration and patient-reported vs. clinician-reported data. The ASES and Constant scores were the most commonly used measures. The UCLA score, which requires clinician inputted data, performed well. The ASES was also found to be responsive and only requires patient-reported data, limiting the burden of reporting on both the patient and the provider. The Constant score performed poorly overall, although the Adjusted Constant showed better results. A better overall understanding of the most appropriate measures to evaluate TSA may lead to a more consistent usage of specific tools, leading to improved capabilities to assess outcomes.

Disclaimer

The authors, their immediate families, and any research foundations 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.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jse.2020.08.019.

References

1. Amstutz HC, Sew How AL, Clarke IC. UCLA anatomic total shoulder arthroplasty. Clin Orthop Relat Res 1981:7-20.
2. Appleby J, Poteliakhoff E, Shah K, Devlin N. Using patient-reported outcome measures to estimate cost-effectiveness of hip replacements in English hospitals. J R Soc Med 2013;106:323-31. https://doi.org/10.1177/0141076813489678
3. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheum 1997;24:799-802.
4. Cleveland O, Piuzzi NS, Strnad G, Brooks P, Hettrich CM, Higuera-Rueda C, et al. Implementing a scientifically valid, cost-effective, and scalable data collection system at point of care: the Cleveland Clinic OME cohort. J Bone Joint Surg Am 2019;101:458-64. https://doi.org/10.2106/JBJS.18.00767

5. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clin Orthop Relat Res 1987:160-4.
6. Davies GM, Watson DJ, Bellamy N. Comparison of the responsiveness and relative effect size of the western Ontario and McMaster Universities Osteoarthritis Index and the short-form Medical Outcomes Study Survey in a randomized, clinical trial of osteoarthritis patients. Arthritis Care Res 1999;12:172-9.
7. Dunlap WP, Cortina JM, Vaslow JB, Burke MJ. Meta-analysis of experiments with matched groups or repeated measures designs. Psych Methods 1996;1:170.
8. Gagnier JJ. Patient reported outcomes in orthopaedics. J Orthop Res 2017;35:2098-108. https://doi.org/10.1002/jor.23604
9. 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 Am 2015;97:e26. https://doi.org/10.2106/JBJS.N.00831
10. Gilbart MK, Gerber C. Comparison of the Subjective Shoulder Value and the Constant score. J Shoulder Elbow Surg 2007;16:717-21. https://doi.org/10.1016/j.jse.2007.02.123
11. Goodman J, Lau BC, Krupp RJ, Getz CL, Feeley BT, Ma CB, et al. Clinical measurements versus patient-reported outcomes: analysis of the American Shoulder and Elbow Surgeons physician assessment in patients undergoing reverse total shoulder arthroplasty. JSES Open Access 2018;2:144-9. https://doi.org/10.1016/j.jses.2018.01.003
12. Hawkins RJ, Thigpen CA. Selection, implementation, and interpretation of patient-centered shoulder and elbow outcomes. J Shoulder Elbow Surg 2018;27:357-62. https://doi.org/10.1016/j.jse.2017.09.022
13. Huang H, Grant JA, Miller BS, Mirza FM, Gagnier JJ. A systematic review of the psychometric properties of patient-reported outcome instruments for use in patients with rotator cuff disease. Am J Sports Med 2015;43:2572-82. https://doi.org/10.1177/0363546514565096
14. Katolik LI, Romeo AA, Cole BJ, Verma NN, Hayden JK, Bach BR. Normalization of the Constant score. J Shoulder Elbow Surg 2005;14:279-85. https://doi.org/10.1016/j.jse.2004.10.009
15. Kirkley A, Griffin S, Dainty K. Scoring systems for the functional assessment of the shoulder. Arthroscopy 2003;19:1109-20. https://doi.org/10.1016/j.arthro.2003.10.030
16. Kujala UM, Jaakkola LH, Koskinen SK, Taimela S, Hurme M, Nelimarkka O. Scoring of patellofemoral disorders. Arthroscopy 1993;9:159-63.
17. Lam ET, Lam CL, Fong DY, Huang WW. Is the SF-12 version 2 Health Survey a valid and equivalent substitute for the SF-36 version 2 Health Survey for the Chinese? J Eval Clin Pract 2013;19:200-8. https://doi.org/10.1111/j.1365-2753.2011.01800.x
18. Matsen FA III, Tang A, Russ SM, Hsu JE. Relationship between patient-reported assessment of shoulder function and objective range-of-motion measurements. J Bone Joint Surg Am 2017;99:417-26. https://doi.org/10.2106/JBJS.16.00556
19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8:336-41. https://doi.org/10.1371/journal.pmed.1000097
20. Nilsdotter AK, Roos EM, Westerlund JP, Roos HP, Lohmander LS. Comparative responsiveness of measures of pain and function after total hip replacement. Arthritis Care Res 2001;45:258-62.
21. Nwachukwu BU, Hamid KS, Bozic KJ. Measuring value in orthopaedic surgery. 2013. JBJS Rev 2013;1:01874474-201311000-00003, https://doi.org/10.2106/JBJS.RVW.M.00067
22. Richards RR, An K-N, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, et al. A standardized method for the assessment of shoulder function. J Shoulder Elbow Surg 1994;3:347-52.
23. Rolfson O, Rothwell A, Sedrakyan A, Chenok KE, Bohm E, Bozic KJ, et al. Use of patient-reported outcomes in the context of different levels of data. J Bone Joint Surg Am 2011;93:66-71. https://doi.org/10.2106/JBJS.K.01021
24. Sciascia AD, Morris BJ, Jacobs CA, Edwards TB. Responsiveness and internal validity of common patient-reported outcome measures following total shoulder arthroplasty. Orthopedics 2017;40:e513-9. https://doi.org/10.3928/01477447-21070327-02
25. Stone AV, Jacobs CA, Luo TD, Meadows MC, Nho SJ, Stubbs AJ, et al. High degree of variability in reporting of clinical and patient-reported outcomes after hip arthroscopy. Am J Sports Med 2018;46:3040-6. https://doi.org/10.1177/0363546517724743
26. Thigpen CA, Shanley E, Momaya AM, Kissenberth MJ, Tolan SJ, Tokish JM, et al. Validity and responsiveness of the Single Alphanumeric Evaluation for shoulder patients. Am J Sports Med 2018;46:3480-5. https://doi.org/10.1177/0363546518807924
27. Unger RZ, Burnham JM, Gammon L, Malempati CS, Jacobs CA, Makhni EC. The responsiveness of patient-reported outcome tools in shoulder surgery is dependent on the underlying pathological condition. Am J Sports Med 2019;47:241-7. https://doi.org/10.1177/0363546517749213
28. Wright RW, Baumgarten KM. Shoulder outcomes measures. J Am Acad Orthop Surg 2010;18:436-44. https://doi.org/10.5435/00124635-201007000-00006