

What Are We Measuring? A Systematic Review of Outcome Measurements Used in Shoulder Surgery



Matthew L. Ashton, B.S., Ian Savage-Elliott, M.D., Caroline Granruth, B.A., and Michael J. O'Brien, M.D.

Purpose: The purpose of this study was to identify the most commonly used outcome measurements following shoulder surgery and to investigate demographic variables related to their use. **Methods:** PubMed and Embase were searched to identify studies in which at least 1 shoulder-specific outcome measurement was used. Exclusion criteria included duplicate studies, review articles, lack of surgical arm, written in a language other than English, or not adult-specific. Additionally, surgeries were subcategorized based on the type of pathology leading to surgery. **Results:** Of the 589 articles identified in the search, 180 met the inclusion criteria. A total of 35 shoulder-specific outcome measurements were reported. The Constant-Murley score (CMS), American Shoulder and Elbow Surgeons Shoulder Score (ASES), Subjective Shoulder Value (SSV), Simple Shoulder Test (SST), and University of California Los Angeles Score (UCLA) were used in more than 10% of the articles. The CMS and SSV were used more commonly together than individually ($P = .0074$). Additionally, the ASES ($P < .00001$) and CMS ($P = .0109$) were associated with the country of origin of the article. The SST was used more frequently in randomized control trials ($P = .0287$). The ASES and DASH were associated with surgeries categorized under the degenerative indication ($P = .001$ and $P = .0146$). Finally, the SSV, ASES and DASH were all found to be significantly paired with surgeries that indicated traumatic pathology ($P = .0061$, $P = .0077$ and $P = .0069$, respectively). **Conclusions:** There is great variability among the outcome measurements currently being used for assessing function following orthopaedic shoulder surgery; however, 5 scoring systems are used more frequently than others. There remains a large discrepancy between the ideal reporting, as noted in the recent literature review, and the current state of outcomes reported at this time. **Clinical Relevance:** By identifying and evaluating the heterogeneity of the reporting and the usage of the performance indicators, these results can guide the standardization of outcome measurements in shoulder surgery and allow for better comparability when assessing outcomes between patients and studies.

Traditionally, shoulder function has been assessed using clinical measures, such as strength, pain and range of motion.¹ However, advances in surgical

technique and prosthesis design and the trend toward patient-reported outcome measures (PROMs) in shoulder surgery have prompted the creation and proliferation of many different clinical-outcome measurements.² Many scoring systems of varying degrees of validity, reliability and responsiveness are currently being reported in the literature concerning shoulder surgery, and more than 25 different PROMs are being used.³⁻⁵ The wide range of outcome measurements makes comparisons among surgeons evaluating various techniques and quality of care extremely difficult. Additionally, widespread evaluation of clinical outcomes is limited. Angst et al.⁴ evaluated the psychometric properties of 9 different outcome measurements commonly used in shoulder surgery. Based on their research, they determined that the Quick Disabilities of Arm, Shoulder & Hand (QuickDASH), the Shoulder Pain and Disability Index (SPADI), the American Shoulder and Elbow Surgeons Score (ASES), and the Constant-Murley Score (CMS) were useful clinical assessment tools, and the DASH, SPADI, and ASES or

From Tulane University School of Medicine, New Orleans, Louisiana, U.S.A. (M.L.A., C.G.), and the Department of Orthopaedic Surgery, Tulane University School of Medicine, New Orleans, Louisiana, U.S.A. (I.S.-E., M.J.O.)

The authors report the following potential conflicts of interest or sources of funding: M.J.O. reports personal fees and other from DePuy, Mitek, and Smith & Nephew, outside the submitted work; personal fees from Stryker, outside the submitted work; is a board or committee member for American Shoulder and Elbow Surgeons, Arthroscopy Association of North America, Association of American Medical Colleges, and the Southern Orthopaedic Association. Full ICMJE author disclosure forms are available for this article online, as [supplementary material](#)

Received July 25, 2019; accepted April 21, 2020.

Address correspondence to Matthew L. Ashton, B.S., 3928 Palmyra St., New Orleans, LA 70119, U.S.A. E-mail: mashton2@tulane.edu

© 2020 THE AUTHORS. Published by Elsevier Inc. on behalf of the Arthroscopy Association of North America. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). 2666-061X/19900

<https://doi.org/10.1016/j.asmr.2020.04.009>

CMS could be recommended for collecting data for future research. Other authors have looked at specific scoring systems, such as the CMS and ASES.⁴⁻⁶ Still others have investigated the efficiency of outcome measurements used exclusively for rotator cuff tears.⁷

Despite these efforts, there is need for a systematic review of literature concerning shoulder surgery that is inclusive of all scoring systems and identifies factors contributing to outcome-measurement use. Similar studies have been conducted recently for both knee and elbow scoring systems, and they yielded important relationships about performance versus established quality standards and cross-cultural validation.^{8,9}

With the drive toward a value-based health care system and the proliferation of scoring systems, standardization of outcome measurements in surgery is of paramount importance. Schmidt et al.¹⁰ performed a systematic review evaluating validity, reliability and responsiveness of 11 PROMs that are applicable to a wide spectrum of shoulder disorders; their intent was to provide recommendations to clinicians and researchers about which PROMs to use. More recently, the American Shoulder and Elbow Surgeons Committee was established; its purpose is to provide recommendations concerning which PROMs to use for patients after shoulder and elbow surgery.³

The purpose of this study was to identify the most commonly used outcome measurements following orthopedic shoulder surgery and to investigate demographic variables related to their use. We hypothesized that there would be a multitude of different scoring systems in use without there being an emphasis on specific factors or variables associated with their use.

Methods

The protocol for this systematic review was performed on the basis of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines.

Search Strategy

A comprehensive systematic review in the PubMed and Embase databases was conducted to obtain articles related to shoulder-specific outcome measurements. The initial PubMed search was conducted on June 18, 2018. We combined key words using Medical Subject Headings (MeSH) terms and free-text entries: ("shoulder"[MeSH Terms] OR "shoulder"[All Fields]) AND outcome[All Fields] AND scores[All Fields] AND value [All Fields]. The initial Embase specific search was conducted on July 20th, 2018 and we used: ("shoulder"[Emtree – major focus exp.] OR "shoulder"[All Fields]) AND outcome[All Fields] AND scores[All Fields] AND value[All Fields]. An updated search was performed on October 10, 2019, for both PubMed and Embase, using the identical search terms. The search

was limited to studies in humans and articles written in English and published after the year 2000.

Eligibility and Exclusion Criteria

In the initial search, 2 authors independently selected articles for inclusion; any disagreement between the authors was further discussed until an agreement was reached. A third author, who had more experience with shoulder surgery, reviewed the final article list created by the first 2 authors to ensure that all articles were shoulder-specific. Exclusion criteria included duplicate studies and any article that was a systematic review or meta-analysis, did not include a surgical component, was not written in English, was not specific to adults (18+), or did not use a shoulder-specific outcome measurement. Generic health-related quality-of-life scoring systems such as the, 36-Item Short Form Health Survey were excluded from our review.

Data Extraction

The 2 authors independently extracted data from all eligible publications and entered them on an electronic spreadsheet. Extracted data included study design, year, surgical procedure, country of origin of lead author, journal of publication, and outcome measurement used. Shortened or altered versions of multiple outcome measurements (i.e., Disabilities of Arm, Shoulder & Hand [DASH] and QuickDASH) were counted as individual outcome measurements.

Data Analysis

Data analysis was conducted by a single author. Outcome measurements that appeared in more than 10% of included articles were selected for analysis. This percentage was used as the cutoff because 5 outcome measurements were used more frequently than others, and all of them were used in more than 10% of the included articles. If any outcome measurements that appeared in less than 10% of the articles were suspected of statistical significance, they were also selected for analysis. Additionally, 2 authors reviewed the types of surgical procedures and types of shoulder conditions (degenerative, instable, traumatic, infectious, autoimmune, or unknown) to assess for any significant scoring system usage associated with either of these variables.

Each outcome measurement was tested alone against a single data measurement using the Fisher exact test. We tested the association between outcome measurement and country of origin, study design, surgical procedure, and year of publication. Additionally, we tested frequency with which 1 outcome measurement was paired with another. For each of the associations, the Fisher exact test was used. Any *P* value of < .05 was considered statistically significant.

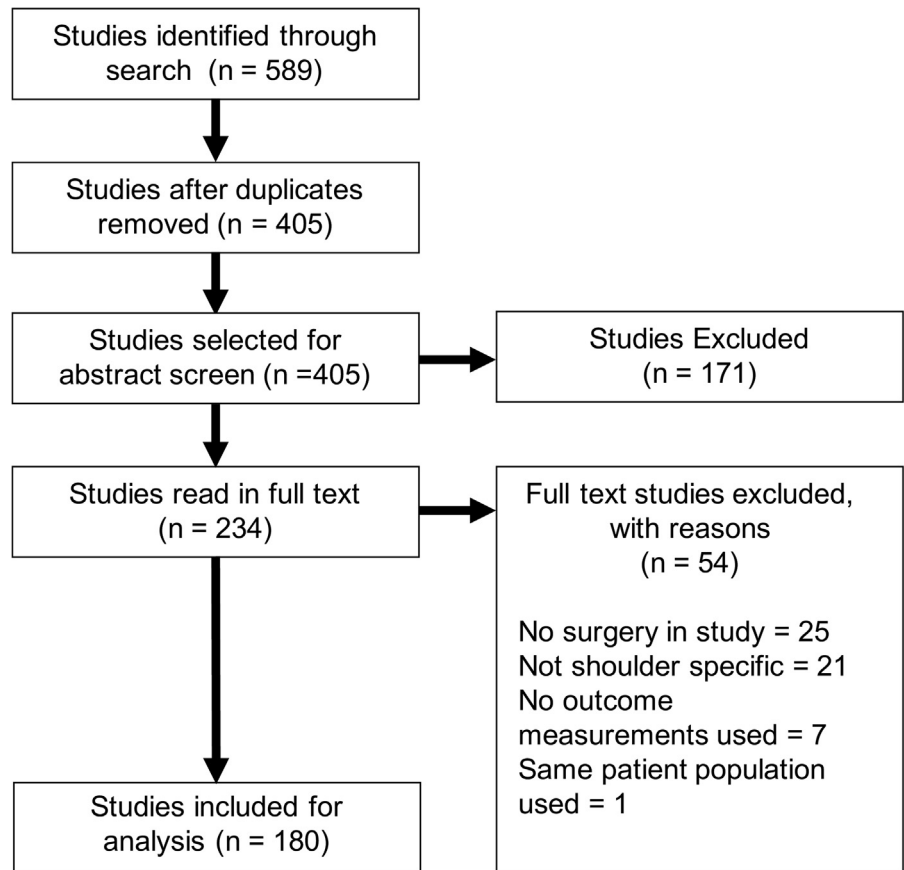


Fig 1. Diagram for study selection. Articles were initially screened by abstract, with subsequent review of remaining articles in full text.

Results

Literature Search

The search strategy yielded a total of 589 publications. After duplicate studies were removed, 405 articles were chosen for assessment of the abstracts. The screening of abstracts excluded 171 studies; the remainder of studies (234) were read in full text. Of those read in full text, 180 met the inclusion criteria and were selected for evaluation (Fig 1).

Scoring Systems

Of the articles evaluated, a total of 35 separate, shoulder-specific outcome measurements were reported. The most frequently reported outcome measurement was the Constant-Murley Score (CMS), used in 52.8% of the included articles. Four additional outcome measurements were found to be included in more than 10% of the articles, including the American Shoulder and Elbow Surgeons Shoulder Score (ASES), the Subjective Shoulder Value (SSV), the Simple Shoulder Test (SST), and the University of California Los Angeles Score (UCLA) (Fig 2). The remaining 30 scores were found in 8% or less of the articles included.

These top 5 outcome measurements were selected for further statistical analysis. We first investigated whether any 2 of these scores were found to be more frequently paired together than others. We found that the use of the CMS and SSV scores together was significant ($P = .0074$). No other statistically significant association between 2 scoring systems was found.

Study of Country of Origin

The top 6 countries publishing results of outcome measurements were the United States, France, Germany, the United Kingdom, Switzerland, and South Korea. Combined, these countries contributed to more than 65% of the total number of studies that were reviewed.

The ASES was found to be significantly paired with articles whose first author was from the United States ($P < 0.00001$). The CMS was found to be significantly paired with articles whose authors were from France, Switzerland or Germany ($P < .00001$, $P = .0109$ and $P = .0318$, respectively). Of the included articles whose first author was from the United States ($n = 59$), the ASES was used in 71.2% (42/59), whereas the CMS was used throughout Europe. No other significant relationships were found.

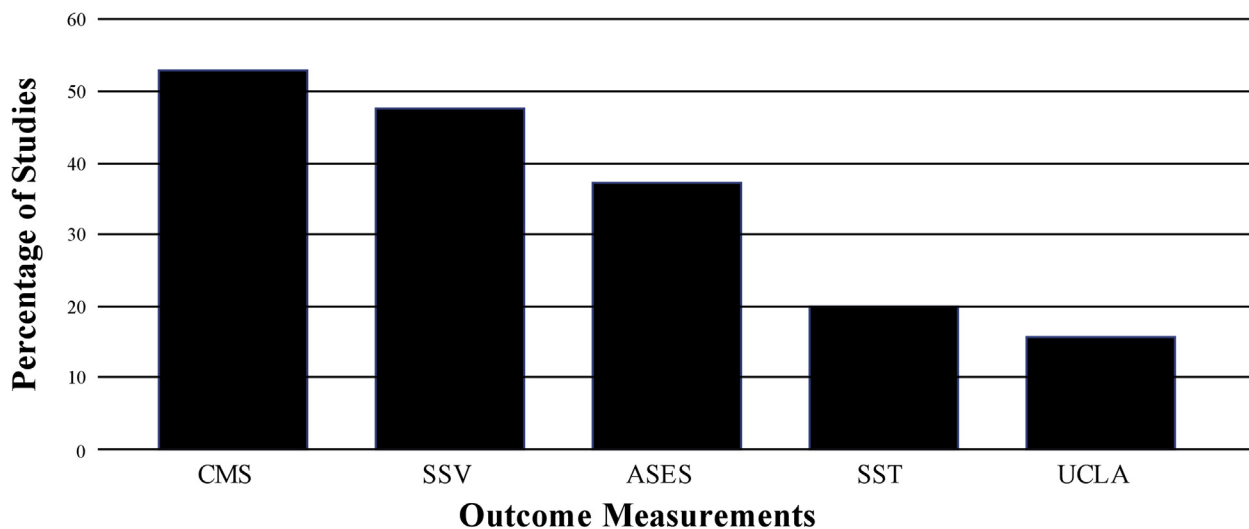


Fig 2. Outcome measurement utilization. These 5 outcome measurements were all used in more than 10% of the articles included in our study. ASES, American Shoulder and Elbow Surgeons Score; CMS, Constant-Murley Score; SST, Simple Shoulder Test; SSV, Subjective Shoulder Value; UCLA, University of California, Los Angeles, Score.

Study Design

Analysis of study design and outcome measurement usage revealed that the SST outcome measurement was used with significant frequency in studies that were randomized controlled trials ($P = .0287$). No other significant associations between study design and outcome measurement were found.

Surgical Procedure

A wide range of surgical procedures was captured in our search. We grouped the types of surgical procedures into various categories based on the indication for surgery—degenerative, instability, trauma, infectious, autoimmune, or unknown.

The ASES and DASH were found to be significantly associated with surgeries categorized as degenerative, which included rotator cuff repair ($P = .001$ and $P = .0146$, respectively). For the instability subgroup, both the Rowe and Western Ontario Shoulder Instability Index (WOSI) were used significantly more often ($P < .00001$ and $P < .00001$). Finally, the SSV, ASES and DASH PROMs were found to be significantly paired with surgeries indicated as trauma ($P = .0061$, $P = .0077$ and $P = .0069$, respectively).

Discussion

The major finding of this study was that 35 shoulder-specific outcome measurements were identified in the shoulder surgery literature, with 5 of these outcome measurements appearing in more than 10% of the articles.

The CMS was the most widely used outcome measurement analyzed, with a usage of 52.8%. This could be attributed to the large number of articles that were generated by a first author whose country of origin was in Europe, where historically, the CMS has been used

and is endorsed by the European Society for Surgery of the Shoulder and the Elbow and the German Society of Shoulder and Elbow Surgeons.^{1,3,4} Vrotsou et al.⁶ performed a systematic review assessing the psychometric properties of the CMS by using the evaluating measures of the Evaluating Measures of Patient-Reported Outcomes (EMPRO) tool. This tool has been proven to be reliable and valid when evaluating condition-specific and generic PROMs.¹¹ Various other shoulder PROMs have also been evaluated using this tool.¹⁰ Vrotsou et al.⁶ used the EMPRO tool to assess various shoulder pathologies, including fractures, arthritis, instability, and frozen shoulder. Their findings suggest that the use of the CMS is advisable for patients with subacromial pathology. However, for other shoulder conditions, the CMS may have the capacity to capture changes over time, but the data were inconclusive. Therefore, Vrotsou et al.⁶ concluded that the CMS should not be considered the gold standard for shoulder evaluation based on the psychometric properties of the CMS when assessed by experts in 7 pathology groups, including osteoarthritis, rheumatoid arthritis and frozen shoulder. The authors did feel, however, that it could generally be applied to subacromial pathology based on its results when analyzing psychometric properties using the EMPRO, a tool designed and validated for standardized assessment of PROMs.⁶

Further research into the applicability of the CMS for generalized shoulder pathology versus only acromioclavicular or subacromial pathology is warranted, given its widespread usage in evaluating the shoulder. It should also be noted that although many of the outcome measurements investigated in this review are PROMs, the CMS is not considered to be so because it includes clinical input to measure strength.⁴

The SSV was the second most widely used outcome measurement in our study (48.3%). One reason for the frequent use of this outcome measurement could be attributed to its ease of use. The SSV was designed with the purpose of providing a simple score reflecting the view of the patient.¹² This score generates a single numeric value that can be easily reported at each patient visit. The SSV was used alone as an outcome measurement only once in our search, whereas each of the other 6 most frequently used outcome measurements included in our study were used alone in multiple articles. The use of the SSV as a stand-alone scoring system versus as an adjunct with other systems requires further investigation.

Of the top 5 outcome measurements used in our study, only the SST was found to be associated with randomized controlled trials. The SST is a practical and standardized measurement tool that assesses patients' shoulder function, and it is understandable by both clinicians and patients.¹³ Schmidt et al.¹⁰ concluded that the SST was recommended for clinical trials, in which responsiveness to change and reproducibility are priorities. The SST also has been reported to have a larger confidence interval than the ASES with regard to shoulder function.⁴ In our study, the SST was found to be used in half of the randomized control trials (50%).

The ASES was employed significantly more in the United States ($P < .00001$), while being used sparingly in other countries. In 2018, the American Shoulder and Elbow Surgeons Value Committee published their recommendations for outcome measurements. They concluded that the ASES was the best available joint-specific outcome measurement to be used for shoulder assessment.³ Additionally, a systematic review performed in 2014 concluded that the ASES is the most valid and reliable outcome measurement for discriminating among patients' or groups' evaluations at 1 point in time.¹⁰ Although both of these studies point to the growing faith in the use of the ASES, our literature review indicates that it has yet to become the standard of care in shoulder assessment.

When the data were analyzed for various pathologies' being associated with specific scoring systems, multiple significant associations were observed. Both the ASES and DASH scoring systems were found to be used significantly more often with degenerative shoulder pathologies ($P = .001$ and $P = .0146$, respectively). Studies of rotator cuff repairs made up a large portion of the degenerative pathology studies, which could explain the significant association with the ASES scoring system because it is used often and has acceptable psychometric performance values for various rotator cuff pathologies.^{4,5} Additionally, degenerative pathologies of the shoulder may have an impact on other joints of the upper extremity, which would make the DASH the optimal choice for a scoring

system because it can incorporate both the elbow and the wrist.

For instability pathologies, both the WOSI and the Rowe scoring systems were found to be used almost exclusively in this category. The American Shoulder and Elbow Surgeons Value Committee recommends the use of the WOSI scoring system for instability, which explains the significant association found in this review.³ The Rowe score was designed in 1978 specifically for assessing stability of the shoulder, which explains its significant association.¹⁴

Three scoring systems were found to be used significantly more often in trauma pathologies: the SSV, ASES and DASH. Again, the DASH assesses function of the entire upper extremity and, therefore, is useful in traumatic injuries that involve more than the shoulder alone.

Limitations

Our study has multiple limitations. This review highlights the heterogeneity of outcomes reporting on shoulder surgery and also subcategorizes each study based on shoulder pathology and outcomes scores used. Classifying the pathologies and the outcomes scores used helps us toward better understanding of the current state of outcomes reporting, but determining the validity of a patient-reported outcome is an extensive process that we believe is outside the scope of this study. Our study is meant to be a helpful commentary on the current status of shoulder-outcomes reporting; however, we acknowledge that it is not an exhaustive summary of each of the clinical outcomes tests studied. Second, although subcategorizing the data further by specific injury and score used would be ideal, we found that these data were incomplete in many studies and, therefore, indication for surgery was used as an alternative measure. Another limitation of this review is the possibility of incomplete reporting. During the data-extraction process, we discovered many discrepancies in the ways in which outcome measurements were reported. Some articles listed all measurements in their abstracts, whereas others listed them only in the full text of the article. We also found variability in the naming of individual outcome measurements in the various articles. For these reasons, our data may not be complete, and that could affect our interpretation of the evidence. Other limitations of our review include the time variation of our data; our search criteria yielded significantly fewer articles from before 2010 than from after that date, making comparison of shoulder-outcome measurements among years difficult. Additionally, our study included a small number of randomized controlled trials (10). Future studies should include more randomized clinical trials in their evaluations of shoulder-specific outcome measurements. Finally, our search was conducted only in the PubMed

and Embase databases; additional studies might have been included if other databases had been searched.

Conclusions

There is a large variability in the outcome measurements that are currently being used for assessing function following orthopaedic shoulder surgery; however, 5 scoring systems are used more frequently than others. There remains a large discrepancy between the ideal reporting, as noted in the recent literature review, and the current state of outcomes reporting at this time.

References

1. Beaton DE, Richards RR. Measuring shoulder function: A cross-sectional comparison of five different questionnaires. *J Bone Joint Surg Am* 1996;78:882-890.
2. Black N. Patient-reported outcome measures could help transform healthcare. *BMJ* 2013;346:f167.
3. Hawkins RJ, Thigpen CA. Selection, implementation, and interpretation of patient-centered shoulder and elbow outcomes. *J Shoulder Elbow Surg* 2018;27:357-362.
4. Angst F, Schwyzer HK, 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 assessment form, (Constant-Murley) Score (CS), Simple Shoulder Test (SST), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), and Western Ontario Shoulder Instability Index (WOSI). *Arthritis Care Res (Hoboken)* 2011;11:S174-S188.
5. Kocher MS, Horan MP, Briggs KK, Richardson TR, O'Holleran J, Hawkins RJ. Reliability, validity, and responsiveness of the American Shoulder and Elbow Surgeons subjective shoulder scale in patients with shoulder instability, rotator cuff disease, and glenohumeral arthritis. *J Bone Joint Surg Am* 2005;87:2006-2011.
6. Vrotsou K, Ávila M, Machón M, et al. Constant-Murley Score: Systematic review and standardized evaluation in different shoulder pathologies. *Qual Life Res* 2018;27:2217-2226.
7. Makhni EC, Hamamoto JT, Higgins JD, et al. How comprehensive and efficient are patient-reported outcomes for rotator cuff tears? *Orthop J Sports Med* 2017;5:232596711769322.
8. Theodoulou A, Bramwell DC, Spiteri AC, Kim SW, Krishnan J. The use of scoring systems in knee arthroplasty: A systematic review of the literature. *J Arthroplasty* 2016;31:2364-2370.
9. Evans JP, Smith CD, Fine NF, et al. Clinical rating systems in elbow research: A systematic review exploring trends and distributions of use. *J Shoulder Elbow Surg* 2018;27:e98-e106.
10. Schmidt S, Ferrer M, González M, et al. Evaluation of shoulder-specific patient-reported outcome measures: A systematic and standardized comparison of available evidence. *J Shoulder Elbow Surg* 2014;23:434-444.
11. Valderas JM, Ferrer M, Garin O, et al. Development of EMPRO: A tool for the standardized assessment of patient-reported outcome measures. *Value Health* 2008;11:700-708.
12. Gilbert MK, Gerber C. Comparison of the subjective shoulder value and the Constant score. *J Shoulder Elbow Surg* 2007;16:717-721.
13. Largacha M, Parsons I, Campbell B, Titelman RM, Smith KL, Matsen F. Deficits in shoulder function and general health associated with sixteen common shoulder diagnoses: A study of 2674 patients. *J Shoulder Elbow Surg* 2006;15:30-39.
14. Rowe CR, Patel D, Southmayd WW. The Bankart procedure: A long-term end-result study. *J Bone Joint Surg Am* 1978;60:1-16.