J Shoulder Elbow Surg (2019) 28, 1341–1346

ELSEVIER

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

General health measures in shoulder surgery: are we powered for success?

David C. Landy, MD, PhD*, Haroutioun Boyadjian, MD, Lewis L. Shi, MD, Michael J. Lee, MD

Department of Orthopedic Surgery, University of Chicago, Chicago, IL, USA

Background: Surgeons, policymakers, and payers increasingly use changes in general health to guide decision-making. It is unknown how such measures are incorporated into shoulder surgery research, how strongly they are associated with changes in shoulder-specific outcomes, and whether they are appropriately powered.

Methods: PubMed was searched for articles reporting shoulder-specific and general health measures after rotator cuff repair and total shoulder arthroplasty. Study characteristics, results at the study group level, reporting of power calculations, and statistical significance were recorded. Meta-regression was employed to describe the association of changes between shoulder-specific and general health measures.

Results: Of 360 identified abstracts, 21 articles with 28 patient groups were included. Only 1 article was published before 2000. There was a strong association of changes between shoulder-specific and general health measures ($r = 0.66$; $P < .001$). Power calculations were mentioned in 33% of studies and based on shoulder-specific measures. Of 20 studies conducting hypothesis tests, 75% reported agreement regarding the statistical significance of shoulder-specific and general health tests. Of 5 discordant studies, 4 found the shoulder-specific measure statistically significant and not the general health measure.

Conclusion: Shoulder surgery research increasingly reports changes in general health measures that are associated with changes in shoulder-specific measures, suggesting that improvements in shoulder symptoms increase quality of life. When disagreement exists, it usually results from the general health measure's not meeting statistical significance, which may simply reflect type II error. Research reporting general health measures should carefully report power considerations to avoid misinterpretation of findings failing to reach statistical significance.

Level of evidence: Level IV; Systematic Review

Keywords: Biostatistics; shoulder; outcome assessment; quality of life

Every year in the United States, hundreds of thousands of rotator cuff repair (RCR) and total shoulder arthroplasty (TSA) operations are performed.[18,28] These operations are almost always elective, with the goal of increasing quality of life through decreased pain and increased function. It is therefore not surprising that studies of such operations would seek to measure changes in general health to help

Institutional Review Board approval was not required for this systematic review.

*Reprint requests: David C. Landy, MD, PhD, 5841 S Maryland Ave, MC3079, Chicago, IL 60637, USA.

E-mail address: David.Landy@UCHospitals.edu (D.C. Landy).

1058-2746/$ - see front matter © 2018 Journal of Shoulder and Elbow Surgery Board of Trustees. All rights reserved.
https://doi.org/10.1016/j.jse.2018.12.012

guide clinical practice.[5,9,21] In addition, policymakers and payers are using such data to decide which interventions to cover.[2,4,26]

It is currently unknown how these general health measures are being incorporated into shoulder surgery research. Furthermore, the association of changes between shoulder-specific and general health measures is not well described. It is unknown whether these efforts are appropriately powered. Given that the hypothesized effect sizes for shoulder-specific measures will be greater than for general health measures, powering studies based on shoulder-specific measures would lead to underpowered evaluations of general health outcomes.[17,31] If inappropriately interpreted, the resultant type II error could underestimate the benefits of shoulder surgery.

We sought to systematically review and to meta-analyze RCR and TSA outcome studies to assess the use of general health measures in shoulder surgery research with 3 specific areas of focus. First, we sought to describe the association of changes between shoulder-specific and general health measures at the study group level. Second, we sought to describe the reporting of power calculations and specifically which type of measure these were based on. Third, we sought to assess the concordance of hypothesis testing between shoulder-specific and general health outcomes from the same studies.

Materials and methods

A systematic review was performed to identify all articles reporting both shoulder-specific and general health measures after RCR and TSA operations using PubMed. All identified abstracts were reviewed, and any article not reporting original research on humans before and after either RCR or TSA was excluded. Articles were then obtained and reviewed. Articles not reporting a mean for the preoperative to postoperative difference in the shoulder-specific and general health measures or a mean for the preoperative and postoperative measures were excluded.

Articles were then systematically reviewed for study characteristics, including publication date, country or origin, and study design. Shoulder-specific and general health outcomes were also recorded. When multiple shoulder-specific measures were reported, the preferred order for recording was American Shoulder and Elbow Surgeons (ASES), Constant, Simple Shoulder Test, and University of California–Los Angeles shoulder score. Other shoulder-specific measures including the Disabilities of the Arm, Shoulder, and Hand score were searched for and coded but not used, given the availability of other measures or limitations in the reporting of that measure. When multiple general health measures were reported, the preferred order for recording was 36-Item Short Form Health Survey (SF-36) physical component summary (PCS) and then 12-Item Short Form Health Survey. Other general health measures were searched for and coded but not used, given the availability of other data or limitations in the reporting of the data.

Next, any details of power calculations, including whether one was reported and, if so, whether it was based on a shoulder-specific or general health measure, were recorded. The statistical significance of hypothesis tests was recorded at the study level for both shoulder-specific and general health measures. When multiple hypotheses were reported, the hypothesis that was most central to the theme of the article was selected. Finally, articles were also reviewed for the reporting of association between shoulder-specific and general health measures at the patient level.

The mean preoperative to postoperative difference was calculated using preoperative and postoperative scores. To allow standardized comparisons across measures, this difference was standardized using the standard deviation of the mean change from preoperative to postoperative, which was estimated using the standard deviations for the preoperative and postoperative means assuming a correlation of 0.25 between preoperative and postoperative measures. Sensitivity analyses were also conducted in which the correlation coefficient for this calculation was allowed to vary between 0 and 0.5 and did not significantly alter the results. When these standard deviations were not reported, they were imputed. To account for differences in precision between studies, meta-analytic weights were calculated as the inverse of the variance of each study group estimate.

Weighted scatter plots were initially used to assess associations between the standardized change in shoulder-specific scores and the standardized change in general health. Weighted linear regression was then used to estimate the magnitude and statistical significance of the association. Furthermore, a weighted version of the Pearson correlation coefficient and an unweighted Spearman rho were calculated to further describe the associations. Given the degree of data transformation, sensitivity analyses were performed using the raw differences in the most commonly reported shoulder-specific and general health measures, the ASES and SF-36.

To assess the potential for publication bias related to the reporting of general health measures, funnel plots were graphed and visually inspected. All analyses were performed using Stata 15.1 (StataCorp, College Station, TX, USA). P values were considered significant at $P < .05$. No attempt was made to adjust for multiple comparisons, although the association between the standardized differences should be viewed as the primary outcome.

Results

Of 360 identified abstracts, 21 articles reporting outcomes for 28 patient groups were included (Fig. 1).[1,3,6-8,10-16,19,20,22-25,27,29,30] Of these articles, only 1 (5%) was published before 2000 and only 7 more (33%) before 2010 (Table I). Most studies were observational and retrospective. In addition, many studies had fewer than 50 patients and reported outcomes for only a single patient group. The mean standardized shoulder-specific change from preoperative to postoperative was 2.2 standard deviations with a range of 2.4 standard deviations. The mean change in general health from these same studies was 1.1 standard deviation with a range of 2.9 standard deviations.

There was a strong association between standardized shoulder-specific and general health measures, with a 1 standard deviation shoulder-specific change predicting a 0.36 standard deviation general health change ($P < .001$; Fig. 2). Both the meta-analytic weighted correlation

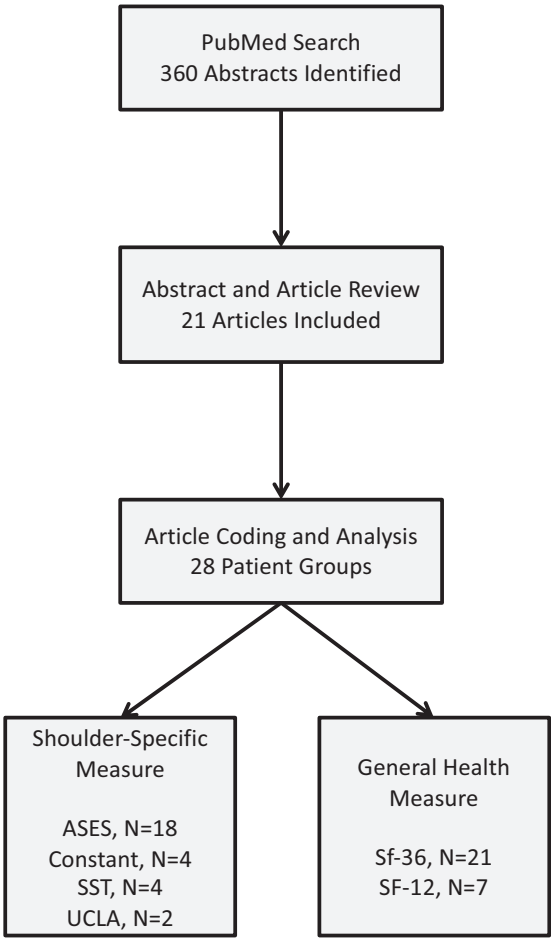

Figure 1 Flow chart showing search strategy and selection process as well as the types of shoulder-specific and general health measures used for each study group. *ASES*, American Shoulder and Elbow Surgeons; *SST*, Simple Shoulder Test; *UCLA*, University of California–Los Angeles shoulder score; *SF-36*, 36-Item Short Form Health Survey; *SF-12*, 12-Item Short Form Health Survey.

coefficient and unweighted Spearman rho were 0.66. The association of unstandardized changes between ASES and SF-36 PCS was only moderate, with a 1-unit change in ASES predicting a 0.13-unit change in SF-36 PCS. The meta-analytic weighted correlation coefficient was 0.52, and the unweighted Spearman rho was 0.34.

Three studies reported associations between shoulder-specific and general health measures, although all did so with respect to preoperative and postoperative measures and not the association of the changes in these measures.[6,15,30] For the specific association of the ASES with the SF-36 PCS, preoperative associations were moderate and ranged from 0.41 to 0.53, and postoperative associations were also moderate and ranged from 0.44 to 0.54.

Power calculations were mentioned in 7 studies (33%). Of these 7 studies, 2 were randomized controlled trials, 3 were prospective observational studies, and 2 were retrospective observational studies. Of these 7 studies, 4 mentioned the specific outcome used for the power calculation, which was always a shoulder-specific measure. Of the 2 retrospective studies, one did include mention of a post hoc power calculation using a general health measure.[20] This study specifically commented on being sufficiently powered for the shoulder-specific measures but underpowered for the general health measures.

Table I Characteristics of 21 included studies

Characteristic	No. (%)
Publication year	
Before 2000	1 (5)
2000-2010	7 (33)
After 2010	13 (62)
Country	
United States	13 (62)
Europe	4 (18)
Canada	2 (10)
South Korea	2 (10)
Study design	
Retrospective observational	11 (52)
Prospective observational	8 (38)
Randomized controlled trial	2 (10)
Study size (No. of patients)	
0-50	11 (52)
51-100	6 (29)
>100	4 (19)
Study groups	
1	11 (52)
2 or more	10 (48)

Of the 20 studies conducting hypothesis tests, 15 (75%) reported agreement regarding the statistical significance of shoulder-specific and general health tests; 11 studies reported that both were statistically significant, and 4 reported that both were not statistically significant (Table II). For the 5 studies with disagreement, 4 found that the shoulder-specific outcome was statistically significant and the general health outcome was not significant. Of the 2 randomized controlled trials, 1 was discordant with the general health outcome failing to reach statistical significance,[24] and neither found the general health outcome to be statistically significant.[22]

A funnel plot was graphed to assess the association between standard error and the standardized change in the general health measures (Fig. 3). There was no obvious trend toward studies with higher standard errors having increasing standardized general health changes.

Discussion

This systematic review of general health measures in shoulder surgery research found that changes in shoulder-specific measures are strongly correlated with changes in general health measures among patients undergoing RCR

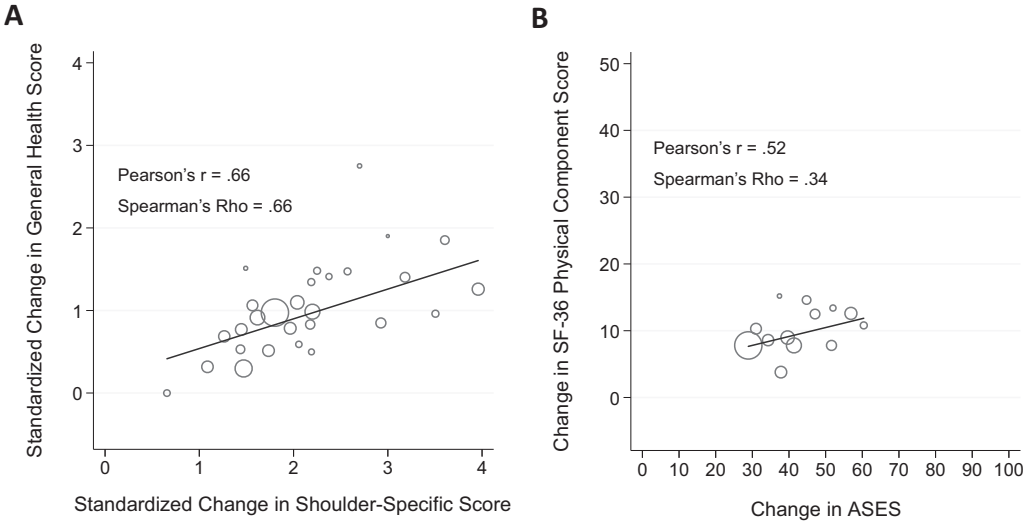

Figure 2 Scatter plot of individual study groups, with observations weighted as the inverse of the variance of the general health estimate, with a linear trend line overlaid, and with correlation coefficients provided. (**A**) The standardized change in general health scores over the standardized change in shoulder-specific scores. (**B**) The change in the 36-Item Short Form Health Survey (SF-36) physical component score over the change in the American Shoulder and Elbow Surgeons (ASES) score.

Table II Association of statistical significance between shoulder-specific and general health hypothesis tests*

		General health	
		Significant	Not significant
Shoulder specific	Significant	11	4
	Not significant	1	4

* Of the 20 studies reporting hypothesis testing for shoulder-specific and general health measures, both the shoulder-specific and general health measure were statistically significant in 11 studies, the shoulder-specific measure and not the general health measure were statistically significant in 4 studies, the general health measure and not the shoulder-specific measure was statistically significant in 1 study, and both the shoulder-specific and general health measures were not statistically significant in 4 studies.

and TSA. However, this review also found that when power calculations are mentioned, studies are powered on the basis of shoulder-specific measures. Furthermore, when disagreement exists, studies most often report that the shoulder-specific measure meets statistical significance, whereas the general health measure fails to meet statistical significance.

A 1 standard deviation change in a shoulder-specific measure was estimated to predict a roughly one-third standard deviation change in a general health measure. As reflected by the strong correlation coefficients, this association was fairly linear across the range examined and consistent across studies. This finding adds to those of other studies that have reported strong associations between shoulder-specific and general health measures both preoperatively and postoperatively at the patient level.[6,15,30] Whereas these strong associations suggest that shoulder disability can have a large impact on quality of life, the strong association between changes in these measures at the study level suggests that changes in shoulder function after surgery translate directly into improved quality of life.

When power calculations were performed prospectively, they were always based on a shoulder-specific measure. This is not surprising, given that shoulder-specific measures would be hypothesized to show a larger effect translating to fewer patients being needed to power such a study compared with a similar study powered on a general health measure. In this study, we found the mean standardized change in shoulder-specific measures to be roughly twice that of general health measures. This practice of selecting a more specific outcome is also consistent with that in other medical specialties.[17] Although this is an expected practice, it does not alter the associated product that general health outcomes from such studies will often be underpowered.[31]

As expected of studies with discordant shoulder-specific and general health outcomes, nearly all found that the general health measure lacked statistical significance. This is further evidence that powering studies based on

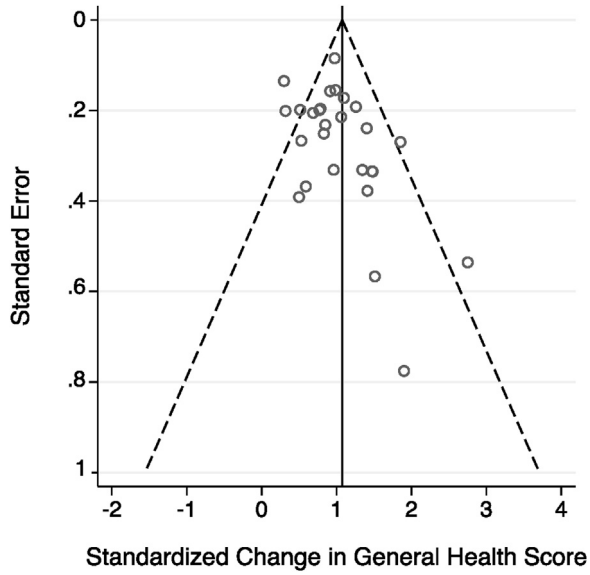

Figure 3 Funnel plot of the standard error by standardized change in general health.

shoulder-specific measures will lead to their being underpowered for comparisons of general health measures. The importance of this issue may even be under-represented in this review, in which many studies were retrospective with sample sizes not based on formal power calculations. Of the 2 randomized controlled trials reviewed, neither found the general health outcome to be statistically significant.

Given the increasing inclusion of general health measures in shoulder surgery research, it is important for readers to be aware of the real danger of type II error with respect to their use as outcomes. Although we are not trying to discourage such investigations, it is important to distinguish secondary analyses of general health measures from appropriately powered a priori hypotheses. It is also important for researchers to be as transparent as possible regarding any power calculations. Especially in prospectively conducted research, reviewers and editors must insist that these details be complete. Certainly, there are limitations to post hoc power analyses; King et al provide a nice example of using such results to clarify how some outcome measures may be sufficiently powered and others not in the same study.[20]

Given the meta-analytic design, this study is vulnerable to publication bias. It is possible that studies with less favorable results went unpublished. In addition, it is possible that despite favorable results for shoulder-specific measures, a study may not have published the general health measures. However, there was no obvious evidence to support these concerns based on the funnel plot. To maximize the data available for analysis and to remove units without specific value, all changes that did require a significant amount of data transformation were standardized. It is reassuring that the estimates from these subsequent analyses were consistent with those from the unstandardized measures, which themselves were consistent with prior literature.

Conclusion

Shoulder surgery studies increasingly report general health measures but infrequently report power calculations. The association of changes between shoulder-specific and general health measures suggests that improvement in shoulder symptoms can significantly improve quality of life. When studies found disagreement in the significance of shoulder-specific and general health measures, most found that the general health measure was not significant, although this may simply reflect type II error as opposed to the true absence of effect. Research reporting general health measures should carefully report power considerations to avoid misinterpretation of findings failing to reach statistical significance. In addition, readers should be familiar with this concern.

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.

References

1. Bhatia S, Greenspoon JA, Horan MP, Warth RJ, Millett PJ. Two-year outcomes after arthroscopic rotator cuff repair in recreational athletes older than 70 years. Am J Sports Med 2015;43:1737-42. https://doi.org/10.1177/0363546515577623
2. Bhat SB, Lazarus M, Getz C, Williams GR Jr, Namdari S. Economic decision model suggests total shoulder arthroplasty is superior to hemiarthroplasty in young patients with end-stage shoulder arthritis. Clin Orthop Relat Res 2016;474:2482-92. https://doi.org/10.1007/s11999-016-4991-0
3. Black AW, Szabo RM, Titelman RM. Treatment of malignant tumors of the proximal humerus with allograft-prosthesis composite reconstruction. J Shoulder Elbow Surg 2007;16:525-33. https://doi.org/10.1016/j.jse.2006.12.006
4. Burwell SM. Setting value-based payment goals—HHS efforts to improve U.S. health care. N Engl J Med 2015;372:897-9. https://doi.org/10.1056/NEJMp1500445
5. Chapman JR, Norvell DC, Hermsmeyer JT, Bransford RJ, DeVine J, McGirt MJ, et al. Evaluating common outcomes for measuring treatment success for chronic low back pain. Spine 2011;36:S54-68. https://doi.org/10.1097/BRS.0b013e31822ef74d
6. Chung SW, Park JS, Kim SH, Shin SH, Oh JH. Quality of life after arthroscopic rotator cuff repair: evaluation using SF-36 and an analysis of affecting clinical factors. Am J Sports Med 2012;40:631-9. https://doi.org/10.1177/0363546511430309

7. Churchill RS. Elective shoulder arthroplasty in patients older than ninety years of age. J Shoulder Elbow Surg 2008;17:376-9. https://doi.org/10.1016/j.jse.2007.09.005
8. Eriksson L, Lindström B, Gard G, Lysholm J. Physiotherapy at a distance: a controlled study of rehabilitation at home after a shoulder joint operation. J Telemed Telecare 2009;15:215-20. https://doi.org/10.1258/jtt.2009.081003
9. Furlong W, Barr RD, Feeny D, Yandow S. Patient-focused measures of functional health status and health-related quality of life in pediatric orthopedics: a case study in measurement selection. Health Qual Life Outcomes 2005;3:3. https://doi.org/10.1186/1477-7525-3-3
10. Gartsman GM, Brinker MR, Khan M. Early effectiveness of arthroscopic repair for full-thickness tears of the rotator cuff: an outcome analysis. J Bone Joint Surg Am 1998;80:33-40.
11. Geervliet P, van den Bekerom M, Spruyt P, Curvers M, Visser C, van Noort A. Short-term results of the global C.A.P. uncemented resurfacing shoulder prosthesis. Orthopedics 2014;37:42-7. https://doi.org/10.3928/01477447-20131219-07
12. Gruson KI, Pillai G, Vanadurongwan B, Parsons BO, Flatow EL. Early clinical results following staged bilateral primary total shoulder arthroplasty. J Shoulder Elbow Surg 2010;19:137-42. https://doi.org/10.1016/j.jse.2009.04.005
13. Gupta AK, Hug K, Berkoff DJ, Boggess BR, Gavigan M, Malley PC, et al. Dermal tissue allograft for the repair of massive irreparable rotator cuff tears. Am J Sports Med 2012;40:141-7. https://doi.org/10.1177/0363546511422795
14. Gupta AK, Hug K, Boggess B, Gavigan M, Toth AP. Massive or 2-tendon rotator cuff tears in active patients with minimal glenohumeral arthritis: clinical and radiographic outcomes of reconstruction using dermal tissue matrix xenograft. Am J Sports Med 2013;41:872-9. https://doi.org/10.1177/0363546512475204
15. Harreld K, Clark R, Downes K, Virani N, Frankle M. Correlation of subjective and objective measures before and after shoulder arthroplasty. Orthopedics 2013;36:808-14. https://doi.org/10.3928/01477447-20130523-29
16. Henn RF 3rd, Tashjian RZ, Kang L, Green A. Patients with workers' compensation claims have worse outcomes after rotator cuff repair. J Bone Joint Surg Am 2008;90:2105-13. https://doi.org/10.2106/JBJS.F.00260
17. Ioannidis DG, Karvouni A, Kouri I, Ioannidis JP. Reporting and interpretation of SF-36 outcomes in randomised trials: systematic review. BMJ 2009;338:a3006. https://doi.org/10.1136/bmj.a3006
18. Jain NB, Higgins LD, Losina E, Collins J, Blazar PE, Katz JN. Epidemiology of musculoskeletal upper extremity ambulatory surgery in the United States. BMC Musculoskelet Disord 2014;15:4. https://doi.org/10.1186/1471-2474-15-4
19. Kemp AL, King JJ, Farmer KW, Wright TW. Reverse total shoulder arthroplasty in wheelchair-dependent patients. J Shoulder Elbow Surg 2016;25:1138-45. https://doi.org/10.1016/j.jse.2015.11.006
20. King JJ, Farmer KW, Struk AM, Wright TW. Uncemented versus cemented humeral stem fixation in reverse shoulder arthroplasty. Int Orthop 2015;39:291-8. https://doi.org/10.1007/s00264-014-2593-6
21. Laucis NC, Hays RD, Bhattacharyya T. Scoring the SF-36 in orthopaedics: a brief guide. J Bone Joint Surg Am 2015;97:1628-34. https://doi.org/10.2106/JBJS.O.00030
22. Lo IK, Litchfield RB, Griffin S, Faber K, Patterson SD, Kirkley A. Quality-of-life outcome following hemiarthroplasty or total shoulder arthroplasty in patients with osteoarthritis. A prospective, randomized trial. J Bone Joint Surg Am 2005;87:2178-85. https://doi.org/10.2106/JBJS.D.02198
23. MacDermid JC, Drosdowech D, Faber K. Responsiveness of self-report scales in patients recovering from rotator cuff surgery. J Shoulder Elbow Surg 2006;15:407-14. https://doi.org/10.1016/j.jse.2005.09.005
24. Moosmayer S, Lund G, Seljom U, Svege I, Hennig T, Tariq R, et al. Comparison between surgery and physiotherapy in the treatment of small and medium-sized tears of the rotator cuff: a randomised controlled study of 103 patients with one-year follow-up. J Bone Joint Surg Br 2010;92:83-91. https://doi.org/10.1302/0301-620X.92B1.22609
25. Mulieri PJ, Holcomb JO, Dunning P, Pliner M, Bogle RK, Pupello D, et al. Is a formal physical therapy program necessary after total shoulder arthroplasty for osteoarthritis? J Shoulder Elbow Surg 2010;19:570-9. https://doi.org/10.1016/j.jse.2009.07.012
26. Odum SM, Hamid N, Van Doren BA, Spector LR. Is there value in retrospective 90-day bundle payment models for shoulder arthroplasty procedures? J Shoulder Elbow Surg 2018;27:e149-54. https://doi.org/10.1016/j.jse.2017.10.008
27. Osti L, Papalia R, Del Buono A, Denaro V, Maffulli N. Comparison of arthroscopic rotator cuff repair in healthy patients over and under 65 years of age. Knee Surg Sports Traumatol Arthrosc 2010;18:1700-6. https://doi.org/10.1007/s00167-010-1081-9
28. Palsis JA, Simpson KN, Matthews JH, Traven S, Eichinger JK, Friedman RJ. Current trends in the use of shoulder arthroplasty in the United States. Orthopedics 2018;41:e416-23. https://doi.org/10.3928/01477447-20180409-05
29. Renfree KJ, Hattrup SJ, Chang YH. Cost utility analysis of reverse total shoulder arthroplasty. J Shoulder Elbow Surg 2013;22:1656-61. https://doi.org/10.1016/j.jse.2013.08.002
30. Yoo JH, Cho NS, Rhee YG. Effect of postoperative repair integrity on health-related quality of life after rotator cuff repair: healed versus retear group. Am J Sports Med 2013;41:2637-44. https://doi.org/10.1177/0363546513499152
31. Zlowodzki M, Bhandari M. Outcome measures and implications for sample-size calculations. J Bone Joint Surg Am 2009;91:35-40. https://doi.org/10.2106/JBJS.H.01602