JSES Open Access 2 (2018) 18–22

Contents lists available at ScienceDirect

JSES Open Access

journal homepage: www.elsevier.com/locate/jses

Quantitative videographic analysis of intraoperative total shoulder arthroplasty is predictive of radiographic implant loosening

Peter Simon, PhD [a,b], Jonathan J. Streit, MD [c], Joseph A. Abboud, MD [d], Mark A. Mighell, MD [b,c], Gerald R. Williams Jr, MD [d], Mark A. Frankle, MD [b,c,*]

[a] Phillip Spiegel Orthopaedic Research Laboratory, Foundation for Orthopaedic Research and Education, Tampa, FL, USA
[b] Department of Orthopaedics and Sport Medicine, University of South Florida, Tampa, FL, USA
[c] Department of Shoulder and Elbow Surgery, Florida Orthopaedic Institute, Tampa, FL, USA
[d] Rothman Institute, Thomas Jefferson University, Philadelphia, PA, USA

ARTICLE INFO

Keywords:
Intraoperative video
Total shoulder arthroplasty
Eccentric glenoid
Walch B2
Video evaluation
Surgical performance

Level of evidence: Basic Science Study, Education Methodology Study, Systems to Improve Surgical Technique

Background: It is believed that both patient and surgeon factors contribute to premature implant loosening. This video study was designed to answer the following questions: Can orthopedic surgeons reliably differentiate between procedures done well and those that will lead to early glenoid failure? Do the difficulty of the operation and the surgeon's performance predict a patient's outcome? Does the presence of a Walch B2 glenoid result in surgery that is evidently more difficult and performed in such a way to suggest early glenoid component failure?

Methods: Eleven upper extremity surgeons blindly graded a set of intraoperative videos of 15 total shoulder arthroplasty patients (grouped by outcome at 2 years). Evaluation questionnaires consisted of questions about the perceived difficulty and the surgeon's performance. Total and partial patient scores were calculated for each video. Higher calculated score would indicate worse postsurgical outcome.

Results: The loosening group had a significantly higher total score ($P = .0057$). Also, patients with B2 glenoids scored significantly higher than patients with other wear type. The analysis of overall procedure performance indicated difference between outcome groups ($P = .0063$).

Conclusion: Our results indicate that surgeons could review surgical videos and differentiate the cases that were difficult or those that were more likely to lead to loosening of the glenoid component. The presence of a B2 glenoid was predictive of difficult surgery. The results of this study should serve as a starting point for surgeons interested in critically evaluating performance and also for those interested in finding ways to maximize patient outcomes after total shoulder arthroplasty.

Total shoulder arthroplasty (TSA) is an end-stage treatment for debilitating shoulder pain resulting from osteoarthritis of the shoulder that provides patients with a predictable and durable outcome in terms of pain relief, function, and quality of life. Although most prosthetic shoulder implants remain solidly fixed, a small percentage show clinical or radiographic signs of early loosening that may result in the need for revision surgery. It is believed that both surgeon factors and patient factors play a role in causing implants to loosen prematurely. Patient factors known to be important include the patient's age, excessive humeral retroversion, rotator cuff quality, and posterior glenoid erosion.[3-7] The presence of a Walch B2 glenoid[13] may be particularly important because often it causes surgical exposure to be difficult, provides inadequate bone stock for component fixation, and likely represents a hostile environment for the implants due to continued asymmetric loading of the new prosthetic joint.[9,12]

Previous study by Birkmeyer et al[1] used video analysis to assess technical skill of the practicing surgeon performing bariatric surgery. The results showed that video-derived peer rating scores can be independently associated with complication rates, operative times, and postoperative readmissions. Although the videographic assessment of technical skills may hold true in the setting of laparoscopic bypass surgery, it is unclear whether intraoperative surgical videos can be used in proficiency assessment in other operative procedures. With pay-for-performance and an increasing demand for objective measures of surgeon and hospital performance, more research in this area has been conducted in recent years.[1,2,8,10,11] However, most of the current research in this area has measured complication risk rather than outcome for patients who do not experience a complication. There is some evidence that surgeons can reliably evaluate the performance of other surgeons using video.[1]

This study was determined to be exempt from review by the Western Institutional Review Board: No. 1-838319-1.

* Corresponding author: Mark A. Frankle, MD, Florida Orthopaedic Institute, 13020 N Telecom Pkwy, Tampa, FL 33637, USA.
E-mail address: mfrankle@floridaortho.com (M.A. Frankle).

https://doi.org/10.1016/j.jses.2017.11.001

We have designed a videographic study of TSA procedures performed by a single, high-volume surgeon to answer the following questions: Can orthopedic surgeons reliably differentiate between surgical procedures that will have good clinical outcome and those that will lead to early glenoid loosening? Do the difficulty of the operation and the surgeon's performance predict a patient's outcome? Does the presence of a Walch B2 glenoid result in surgery that is evidently more difficult and that is more likely to lead to early glenoid loosening?

Materials and methods

For the purposes of this pilot study, we used a prospectively collected database of all patients undergoing TSA by a single, high-volume upper extremity surgeon at our institution that consists of preoperative and postoperative imaging, range of motion, and outcome scores as well as intraoperative video of the entire case. The available pool for our intraoperative video selection consisted of a total of 344 subjects who underwent TSA between 2004 and 2011. Each video was captured intraoperatively from the initial surgical approach to the closure of the subscapularis. Inclusion criteria for initial patient population pool were defined as diagnosis of primary glenohumeral osteoarthritis, existence of preoperative computed tomography and radiographic assessment and intraoperative video, and a minimum of 2 years of clinical and radiographic follow-up. All subjects with a history of prior arthroplasty as well as any diagnosis other than primary glenohumeral osteoarthritis were excluded from the study. The Foundation and Turon Total Shoulder System (DJO Surgical, Austin, TX, USA) was used for all subjects.

The preoperative and 2-year postoperative data regarding age, sex, American Shoulder and Elbow Surgeons (ASES) score, Simple Shoulder Test, visual analog scale pain score, and range of motion were also available for each patient. Based on the collected 2-year minimum follow-up outcome and imaging data, 16 intraoperative videos were selected from the available patient pool as follows: 8 videos were randomly selected from the pool of patients with good outcomes (sorted by the highest values for preoperative forward flexion, abduction, ASES, and function scores, lowest for pain score) at 2 years who also had no evidence of radiographic implant loosening; 8 additional randomly selected videos from 8 patients who developed glenoid component loosening within the first 2 years of follow-up were also included for the analysis (Table I). An evaluation of initial postoperative and most recent postoperative (minimum 2 years) radiographs (anteroposterior, Grashey) was performed by the senior author (M.A.F.). Glenoid component was considered loosened when the most recent postoperative radiographs presented the radiographic evidence of gross migration of the polyethylene component, reflected by grade 4 or 5 on the Lazarus scale. One video was excluded later in the process because of its low visual quality reported by the evaluators.

The videos were viewed by the evaluators on their home computers using an organized online repository, and ratings were entered into a database immediately after viewing. Audio record was removed to ensure unbiased evaluation of each procedure. Evaluators were blinded to the demographics of the patients, preoperative and postoperative radiographs, and ultimate outcomes of the patients undergoing surgery.

We developed a questionnaire designed to evaluate the opinion of observers blinded to the patient's radiographs and outcome regarding the difficulty of the operation as well as the performance of the procedures. In total, each evaluator (11 evaluators) would score a set of 24 questions for the 15 intraoperative videos. A scale was designed to indicate the opinion of the grader on the perceived difficulty and surgical performance rather the overall importance of the surgical step. Questions about the perceived difficulty of the procedure had 3 levels on a Likert scale (1, easy; 2, average; 3, difficult), and questions about the performance of the procedure had 5 levels (1, excellent; 2, good; 3, average; 4, below average; 5, poor). Questions were sorted into 12 groups representing stages of TSA surgery, such as surgical approach, glenoid exposure, and subscapularis repair (Table II). A total patient score would be calculated for every patient's intraoperative video evaluated by each surgeon as the sum of individual scores per each question. In addition, scores assigned to questions about surgical difficulty as well as the scores assigned to the questions about surgical performance were summed and individually analyzed. An ideal case of easy surgery (12 points) with an excellent performance (12 points) would yield 24 points, whereas the worst-case scenario of difficult case (36 points) with poor performance at every stage (60 points) would yield a score of 96 points.

Table I
Basic demographics and outcome data between groups

Parameter		Good outcome group	Glenoid loosening group	Mann-Whitney *U* test
Follow-up time	Mean	32.9	37.3	.536
	STD	6.9	10.4	
Age	Mean	63.9	67.3	.694
	STD	10.8	5.9	
Gender	Men	3	5	N/A
	Woman	4	3	
ASES pain	Mean	47.9	25.7	.072
	STD	3.9	21.5	
ASES function	Mean	45.2	30.2	.094
	STD	5.1	19.9	
FF	Mean	174.3	125.7	**.029**
	STD	11.3	62.7	
AB	Mean	157.1	120	.094
	STD	32.5	62.2	
ER	Mean	55.7	60	.955
	STD	40.4	23.1	
IR	Mean	5.3	3.3	.121
	STD	2.1	1.9	
SST total	Mean	7.3	4.9	.535
	STD	5.3	4.5	
ASES total	Mean	92.4	55.9	**.014**
	STD	8.8	30.8	
Pain	Mean	0.4	4.9	.072
	STD	0.8	4.3	
Function	Mean	9.6	4.4	**.006**
	STD	0.5	3.6	

ASES, American Shoulder and Elbow Surgeons; *FF*, forward flexion; *AB*, abduction; *ER* external rotation; *IR* internal rotation; *SST*, Simple Shoulder Test; *STD*, standard deviation; *N/A*, not applicable.
Boldface text indicates p-values that reached statistical significance.

Statistical analysis

A Mann-Whitney *U* test was used to evaluate differences in clinical outcome between good outcome (n = 7) and loosening groups (n = 8). An analysis of variance and Fisher post hoc test were used to compare individual scores between outcome groups as well as between wear patterns, classified radiographically by 3 orthopedic surgeons using the Walch classification.[13] Interclass correlation coefficient was used to calculate rater association, and analysis of variance and Pearson correlation were used to estimate rater bias. The α was set at .05.

Results

The patient groups selected for loosening (n = 7) and nonloosening (n = 8) were not different in terms of age (loosening, mean 63.9 years; nonloosening, mean 67.3 years; $P = .69$) and demonstrated predictably different results in terms of clinical outcomes. Patients with loosening demonstrated lower ASES scores (55.9 vs. 92.4; $P = .01$),

Table II
Questionnaire sorted by the stage of TSA surgery

Group	Stage of TSA surgery	Question about difficulty	Question about performance
1	Approach	Ease of approach	Adequacy of approach
2	Glenoid exposure	Ease of glenoid exposure	Adequacy of glenoid exposure
3	Humeral exposure	Ease of humeral exposure	Adequacy of humeral exposure
4	Glenoid component placement	Ease of glenoid component placement	Adequacy of glenoid component placement
5	Humeral placement	Ease of humeral placement	Adequacy of humeral placement
6	Glenoid sizing	Difficulty in glenoid sizing	Adequacy of glenoid sizing
7	Humeral sizing	Difficulty of humeral sizing	Adequacy of humeral sizing
8	Subscapularis repair	Difficulty of subscapularis repair	Adequacy of subscapularis repair
9	Glenoid ROM evaluation	Ease of intraoperative ROM evaluation	Adequacy of intraoperative ROM evaluation
10	Glenoid tissue balance	Difficulty in soft tissue balance	Adequacy of soft tissue balance
11	Glenoid surface preparation	Ease of appropriate surface preparation	Adequacy of surface preparation
12	Overall	Overall ease of surgery	Overall adequacy of surgery

TSA, total shoulder arthroplasty; *ROM*, range of motion.

Table III
Estimated values of rater association between fellowship trainees (Pearson correlation)

	Trainee 1	Trainee 2	Trainee 3	Trainee 4	Trainee 5	Trainee 6	**Trainee 7**
Trainee 1		**0.050**	**0.005**	0.067	**0.001**	0.052	**0.399**
Trainee 2			0.362	0.894	0.199	0.982	**0.006**
Trainee 3				0.297	0.705	0.351	**0.000**
Trainee 4					0.156	0.911	**0.008**
Trainee 5						0.191	**0.000**
Trainee 6							**0.006**
Trainee 7							

Boldface text indicates p-values that reached statistical significance.

lower forward flexion (125.7 vs. 174.3; $P = .03$), and trends toward greater visual analog scale pain scores (4.9 vs. 0.4; $P = .07$) and less active abduction (120 vs. 157.1; $P = .09$) at the time of 2-year follow-up (Table I).

Eleven evaluators at different levels of training and experience evaluated the videos and completed the questionnaire. The evaluators consisted of 4 upper extremity surgeons (10+ years in practice) and 7 upper extremity surgery fellowship trainees. Each video was evaluated only once by each rater, and all evaluators watched and rated all 15 videos included in this study. Consistency of grading was evaluated using an interclass correlation coefficient calculated for the final grade. Average value for interclass correlation was 0.813 for all graders. Two graders repeated evaluation after a washout period of 2 weeks. An average interclass correlation coefficient for repeated grades was 0.870.

When the level of experience was considered, there was no significant difference in the total score between senior graders ($P = .336$). On the other hand, the total grades assigned by the fellowship trainees showed variations with statistical significance ($P < .0001$). Two fellowship trainees deviated from the rest of the group (Table III). For the rest of the analysis, the grades assigned by trainee 1 and trainee 7 were excluded.

A higher risk of loosening was predicted by a higher total video score ($P = .007$), which took into account both the difficulty of the procedure and the performance of the surgeon (Fig. 1). Subanalysis of the difficulty and performance scores in relation to loosening revealed loosening to be directly related to performance ($P = .005$) but not difficulty ($P = .09$) (Fig. 2).

An effect of glenoid wear pattern has been shown to be a significant factor ($P < .0001$). Surgery performed in patients with B2 glenoids resulted in higher total video scores compared with all other glenoid types (A1, $P = .002$; A2, $P = .013$; B1, $P = .001$; C, $P < .0001$) (Fig. 3). Subanalysis found that surgery for patients with B2 glenoids was more difficult compared with all other types except for A2 (A1, $P = .02$; A2, $P = .134$; B1, $P = .003$; C, $P = .031$), and surgery performance was rated as less adequate for B2 compared with all other types (A1, $P = .003$; A2, $P = .011$; B1, $P = .004$; C, $P < .0001$) (Fig. 4).

Discussion

In this pilot study of TSA using video, 11 orthopedic surgeons evaluated the difficulty of surgery and estimated the likelihood of development of early glenoid loosening based on the performed procedure with good agreement. The presence of a Walch B2 glenoid resulted in surgery that was more difficult and that was rated as more likely to lead to glenoid component loosening. This is the first study in orthopedics to relate the surgeon's performance to the patient's outcomes and the only study that we know of in all of medicine to relate the surgeon's performance and degree of difficulty to outcomes rather than to complications.

The finding of greatest significance in this study was that a Walch B2 glenoid results in surgery that is very difficult to perform well. In addition to the known problems relating to the available bone stock for fixation, such a high level of difficulty may help explain

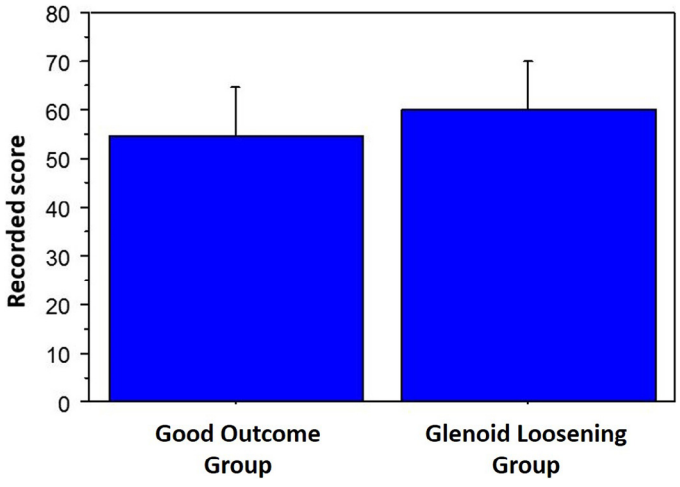

Figure 1 Total score: comparison between outcome groups.

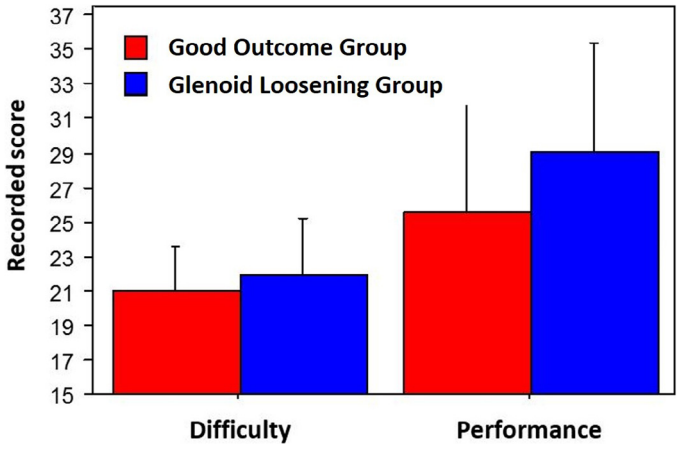

Figure 2 Difficulty and performance individual scores: comparison between outcome groups. Difficulty: good vs. loosening, $P = .0848$. Performance: good vs. loosening, $P = .0063$.

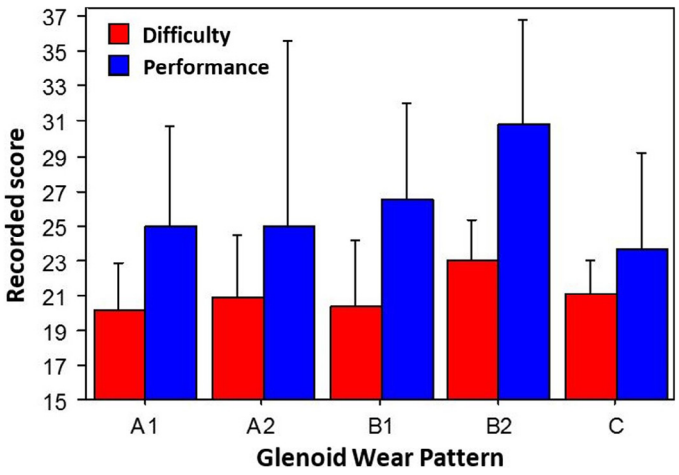

Figure 4 Difficulty and performance individual scores: comparison between wear patterns. Difficulty: B2 vs. A1, $P = .0003$; B2 vs. A2, $P = .0726$; B2 vs. B1, $P = .0010$; B2 vs. C, $P = .0410$. Performance: B2 vs. A1, $P = .0006$; B2 vs. A2, $P = .0238$; B2 vs. B1, $P = .0097$; B2 vs. C, $P = .0003$.

why TSA often produces inferior outcomes for these patients even when it is performed by experienced surgeons.[9] This finding may influence surgical practice by encouraging earlier intervention in patients with osteoarthritis who are developing a B2 glenoid as this surgery may be somewhat easier to perform well. However, this is a pilot study performed on a limited number of operations. Further analysis is required to evaluate other indications and patient-related factors. In addition, surgeons may consider alternative strategies to treat patients with advanced disease, such as personalized instrumentation systems or reverse shoulder arthroplasty, in an effort to reduce technical difficulty and to provide a more predictable surgical performance.

Surgeons in this study were able to discern whether surgery was being performed with ease and with technical skill. This presents an interesting question about an important aspect of the surgeon's experience that is rarely considered: Does a surgeon know when a procedure is not going well, and does he or she consciously make corrections based on this "gut feeling"? If so, how does a surgeon learn to recognize such difficulties and adapt, rather than pushing ahead with an operation that may not provide the patient with a good outcome? We think it is likely that the surgeon's experience plays a role in this type of ability. Such behavior is difficult to quantify, but studies such as ours should provide a basis for discussion among surgeons and may lead to more open dialogue in a field that traditionally has valued strength and resilience over the realization of shortcomings in the operating room. Patient care will undoubtedly be improved when surgeons are able to honestly assess performance "in real time" and adapt as necessary.

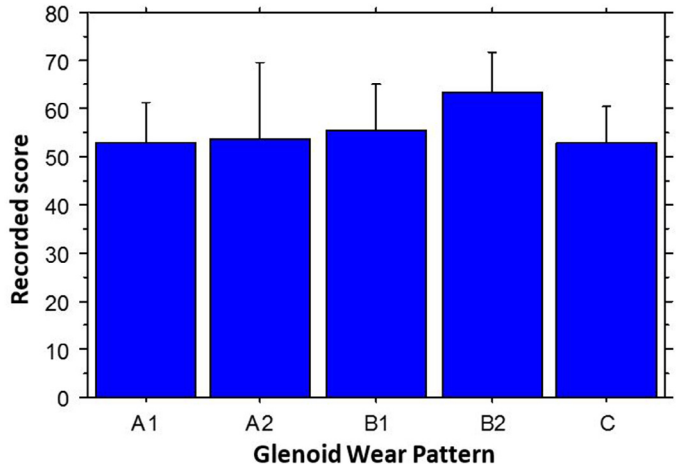

Figure 3 Total score: comparison between wear patterns.

This pilot study had some limitations that must be addressed and understood. First, the surgical videos assessed in this study were not part of a consecutive series, and they were not randomly selected. We specifically chose patients for the loosening group based on a poor outcome. This may be considered a deviation from the standard practice of random selection. However, we were attempting to study reasons for failure, which is, fortunately, not a common occurrence after TSA. The distribution of glenoid types included in the study, although not random, was not specifically selected to include each different type. Second, a single surgeon performed all operations but had a variation of surgical assistants. This introduces some variation in the surgical technique, although all surgeons would be expected to proceed in a similar manner based on training with this single surgeon. Finally, we used a small number of patients in our study, but we were still able to find a difference in ratings of surgical performance, indicating adequate power and a large effect size.

Conclusion

We have evaluated surgeon ratings of TSA surgical cases using video to study the effect of surgical performance on outcome, and our results indicate that surgery that was difficult or performed less adequately led more often to early loosening of the components. The presence of a B2 glenoid was predictive of difficult surgery. The results of this pilot study should serve as a starting point for surgeons interested in critically evaluating performance and also for those interested in finding ways to maximize patient outcomes after TSA.

Disclaimers

DJO Surgical has provided general research support to the Foundation for Orthopaedic Research and Education.

Peter Simon reports that the research foundation with which he is affiliated received research support from DJO Surgical, a

designer and manufacturer of orthopedic surgical products related to the subject of this work.

Jonathan J. Streit, his immediate family, and any research foundation with which he is affiliated has not received any financial payments or other benefits from any commercial entity related to the subject of this article.

Joseph A. Abboud receives consulting fees from Integra Life Sciences, DePuy Synthes, Tornier, and DJO Global and royalties from Integra Life Sciences and Lippincott Williams & Wilkins.

Mark A. Mighell receives consulting fees and honoraria for educational services from DJO Surgical and Stryker.

Gerald R. Williams Jr receives royalties from DePuy, Lippincott Williams & Wilkins, IMDS/Cleveland Clinic, and DJO Surgical; he receives research support from Tornier, DePuy, and Synthasome; and he has stock ownership in In-vivo Therapeutics, ForMD, CrossCurrent Business Analytics, OBERD, and Force Therapeutics.

Mark A. Frankle receives royalties from and is a paid consultant for DJO Surgical.

References

1. Birkmeyer JD, Finks JF, O'Reilly A, Oerline M, Carlin AM, Nunn AR, et al. Surgical skill and complication rates after bariatric surgery. N Engl J Med 2013;369:1434-42. http://dx.doi.org/10.1056/NEJMsa1300625
2. Birkmeyer JD, Stukel TA, Siewers AE, Goodney PP, Wennberg DE, Lucas FL. Surgeon volume and operative mortality in the United States. N Engl J Med 2003;349:2117-27. http://dx.doi.org/10.1056/NEJMsa035205
3. Boileau P, Avidor C, Krishnan SG, Walch G, Kempf J-F, Molé D. Cemented polyethylene versus uncemented metal-backed glenoid components in total shoulder arthroplasty: a prospective, double-blind, randomized study. J Shoulder Elbow Surg 2002;11:351-9. http://dx.doi.org/10.1067/mse.2002.125807
4. Farron A, Terrier A, Büchler P. Risks of loosening of a prosthetic glenoid implanted in retroversion. J Shoulder Elbow Surg 2006;15:521-6. http://dx.doi.org/10.1016/j.jse.2005.10.003
5. Gerber C, Costouros JG, Sukthankar A, Fucentese SF. Static posterior humeral head subluxation and total shoulder arthroplasty. J Shoulder Elbow Surg 2009;18:505-10. http://dx.doi.org/10.1016/j.jse.2009.03.003
6. Godenèche A, Boileau P, Favard L, Huec J-CL, Lévigne C, Nové-Josserand L, et al. Prosthetic replacement in the treatment of osteoarthritis of the shoulder: early results of 268 cases. J Shoulder Elbow Surg 2002;11:11-8. http://dx.doi.org/10.1067/mse.2002.120140
7. Hasan SS, Leith JM, Campbell B, Kapil R, Smith KL, Matsen FA III. Characteristics of unsatisfactory shoulder arthroplasties. J Shoulder Elbow Surg 2002;11:431-41. http://dx.doi.org/10.1067/mse.2002.125806
8. Hervey SL, Purves HR, Guller U, Toth AP, Vail TP, Pietrobon R. Provider volume of total knee arthroplasties and patient outcomes in the HCUP-nationwide inpatient sample. J Bone Joint Surg Am 2003;85-A:1775-83.
9. Hussey MM, Steen BM, Cusick MC, Cox JL, Marberry ST, Simon P, et al. The effects of glenoid wear patterns on patients with osteoarthritis in total shoulder arthroplasty: an assessment of outcomes and value. J Shoulder Elbow Surg 2015;24:682-90. http://dx.doi.org/10.1016/j.jse.2014.09.043
10. Katz JN, Barrett J, Mahomed NN, Baron JA, Wright RJ, Losina E. Association between hospital and surgeon procedure volume and the outcomes of total knee replacement. J Bone Joint Surg 2004;86:1909-16.
11. Lau RL, Perruccio AV, Gandhi R, Mahomed NN. The role of surgeon volume on patient outcome in total knee arthroplasty: a systematic review of the literature. BMC Musculoskelet Disord 2012;13:250. http://dx.doi.org/10.1186/1471-2474-13-250
12. Steen BM, Cabezas AF, Santoni BG, Hussey M, Cusick MC, Kumar AG, et al. Outcome and value of reverse shoulder arthroplasty for treatment of glenohumeral osteoarthritis: a matched cohort. J Shoulder Elbow Surg 2015;24:1433-41. http://dx.doi.org/10.1016/j.jse.2015.01.005
13. Walch G, Badet R, Boulahia A, Khoury A. Morphologic study of the glenoid in primary glenohumeral osteoarthritis. J Arthroplasty 1999;14:756-60.