Contents lists available at ScienceDirect

JSES Open Access

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

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]

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),

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.

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.

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),