A Correlation of Five Commonly Used Clinical Metrics to Measure Outcomes in Shoulder Arthroplasty

Flurin, PH; Wright, T; Zuckerfman, J; Johnson, D; Christensen, M; Roche, C;
1Bordeaux-Mernignac Clinic, FR; 2Univ. of Florida Dept. of Ortho., Gainesville, FL; 3Hosp. for Joint Diseases, NY; 4Exactech, Gainesville, FL; (352) 377-1140. Fax: (352) 378-2617. chris.roche@exact.com

Introduction
Standardizing the tools and methods by which healthcare professionals collect clinical outcomes is a critical component of evidence-based medicine. Numerous outcome measurement scores are available to evaluate the success of treatment of patients with debilitating conditions in the shoulder. The most commonly used scoring systems are the Simple Shoulder Test (SST), the UCLA Score, the American Shoulder and Elbow Surgery Score (ASES), the Constant-Murley (Constant), the Shoulder Pain and Disability Index (SPADI), the Disabilities of the Arm, Shoulder, and Hand (DASH), the Rowe Score, and the Oxford Shoulder Score. While each metric attempts to rate the quality of care, each varies by the method in which it gauges the success of treatment (based upon the restoration of function, motion, and strength and by the reduction of pain). Additionally, each metric varies in how it distributes and weighs subjective patient responses and clinical observations from objective clinician/independent examiner assessment measurements.

In the US, the five most commonly used scoring systems for shoulder arthroplasty are the SST, UCLA, ASES, Constant, and SPADI clinical metrics. The SST score is derived from a series of 12 Yes /No questions that measure the patient’s ability to carry out activities of daily living; 12 is the highest/best score. The UCLA score is derived from a series of 5 questions that evaluate pain, satisfaction of treatment, and restoration of function, strength, and motion; 35 is the highest/best score. The ASES score is derived from a series of 11 questions that evaluates pain (50%) and restoration of function (50%); 100 is the highest/best score. The Constant score is derived from a series of 23 questions that evaluates pain (15%), restoration of function (20%), range of motion (40%), and strength/power (25%); 100 is the highest/best score. The SPADI score is derived from a series of 13 questions that evaluates pain and restoration of function; 130 is the highest score and 0 is the best score. The pre-op and post-op outcomes data was collected and scored using the 5 aforementioned metrics on 45 patients who received a primary reverse shoulder for the treatment of cuff tear arthropathy (CTA). The pre-op and post-op scores were normalized, correlated, and compared to gain a better understanding of the relationship between the metrics.

Methods
45 Patients (age = 73.9 ± 5.9 yrs; 34 females; 34 right shoulders) received a primary Equinoxe reverse shoulder (Exactech, Inc; 30 38mm glenospheres and 15 42mm glenospheres) by the senior author (PHF) through the delta-pectoral approach between May 2007 and June 2010 for treatment of CTA. These patients were evaluated and scored pre-operatively and at latest follow-up using the SST, UCLA, ASES, Constant and SPADI scoring metrics; the average follow-up for all patients was 25.3 ± 10.3 months. A Student’s two-tailed, paired t-test was used to identify differences in pre-operative and post-operative results, where p<0.05 denoted a significant difference. In order to compare the results on the same scale, all 5 metrics were normalized on a 100 point scale. The normalized scores were then correlated to one another to gain a better understand the relationship between the metrics.

Results
The average pre-op and post-op outcomes scores are presented in Table 1. The average pre-op and post-op objective clinician assessment for active abduction, active forward flexion, and active external rotation are presented in Table 2. No instances of instability were reported; the only complications observed were 4 instances of scapular notching (3 Grade 1 and 1 Grade 2) for a rate of 8.9%. Table 3 presents the method by which each metric was normalized to a 100 point scale and also presents the normalized pre-op and post-op values for each scoring metric. Table 4 presents the correlation between the scoring systems.

Comparing the normalized pre-op scores, the SST score was significantly different than UCLA (p<0.001), ASES (p<0.001), Constant (p<0.001), and SPADI (p<0.001). The UCLA score was significantly different than the ASES (p<0.001) and Constant (p<0.001). The ASES score was significantly different than the Constant score (p=0.027). Comparing the normalized post-op scores, the SST score was significantly different than UCLA (p=0.0127), ASES (p=0.0139), Constant (p<0.001), and SPADI (p=0.013). The UCLA score was significantly different than the Constant (p<0.001) and SPADI (p=0.022). The ASES score was significantly different than the Constant (p<0.001). The Constant score was significantly different than the SPADI (p<0.001).

Discussion and Conclusions
The results of this study demonstrate that the 5 clinical metrics utilized in this outcomes study to measure the short term results of a reverse shoulder prosthesis are all very highly correlated (>94%). All 5 metrics demonstrated significant improvements in treatment using the Equinoxe reverse shoulder at a mean follow-up of 25.3 months. Despite this agreement and high degree of correlation, there was a wide difference in the percent increase and mean values between the normalized pre-op and post-op scores. These differences likely arise from the different weights within the scoring systems (e.g. reduction of pain, restoration of function, motion and strength) and from how subjective measures and objective measurements are distributed. Additional work is required to better understand how these differences in weights and methods within systems influence the evaluation and scoring of patient treatment.

Significance
This outcomes study reports on the relationship between 5 commonly used metrics that clinicians use to quantify the treatment of shoulder arthroplasty. The normalized results demonstrate that each metric is highly correlated to each other, though statistically different.