Association between the Goutallier Classification and Objective, Image-Based Measures of Fatty Infiltration in the Rotator Cuff

Josie Elwell[1], Hamidreza Rajabzadeh-Oghaz [1], Gregory Spangenberg[1], William R. Aibinder[2], Bradley Schoch[3], Bruno Gobbato[4], Scott W. Trenhaile[5] Christopher P. Roche[1]
[1]Exactech Inc., Gainesville, FL, USA, [2]University of Michigan, Ann Arbor, MI, USA [3]Mayo Clinic, Jacksonville, FL, US, [4]R. José Emmendoerfer, Jaraguá do Sul, SC, Brazil, [5]OrthoIllinois, Rockford, IL, USA

Disclosures: J. Elwell: 3A; Exactech, Inc. **H. Rajabzadeh-Oghaz:** 3A; Exactech, Inc. **G. Spangenberg:** 3A; Exactech, Inc. **W. Aibinder:** 3B; Exactech, Inc. **B. Schoch:** 3B; Exactech, Inc. **B. Gobbato:** 3B; Exactech, Inc. **S. Trenhaile:** 3B; Exactech, Inc. **C. Roche:** 3A; Exactech, Inc..

INTRODUCTION: Fatty infiltration (FI) of the rotator cuff, whether due to aging or tendinous tears, may ultimately impact the outcomes of potential treatments for degenerative pathologies that necessitate joint replacement and thus is an important consideration in treatment decision-making. The most common method to assess FI of the rotator cuff clinically is the Goutallier classification, a subjective grading scale from 0-4 based on the relative ratio of fat to muscle. However, this scale is coarsely discretized and subject to intra- and inter-observer error. Image processing techniques to objectively quantify FI on continuous scale have been proposed, where the pixels/voxels within 2D or 3D region of interest (ROI) are classified into fat or muscle deterministically based on pre-defined grey-scale value ranges. Radiomics is a field where many HU-based features of a ROI are objectively quantified, however it is unknown if any of these features correlate to the amount of FI in a muscle. Therefore, the objective of this pilot study was to quantify radiomic features of rotator cuff muscles, specifically the supraspinatus (SUP) and the subscapularis (SUB), and determine if they are correlated to surgeon-assessed FI.

METHODS: A multi-center, international database of patients undergoing shoulder arthroplasty with a single implant system was queried retrospectively for patients with available pre-operative CT scans and Goutallier classification of the SUP and SUB. Data were collected via standardized, IRB-approved forms. Goutallier classification of 0 was considered no FI, 1-2 was low FI, and 3-4 was high FI. Accordingly, 100 CT scans were randomly selected while enforcing the following sample sizes: 25 patients with high FI in both the SUP and SUB, 25 with high SUP FI/low SUB FI, 25 with low SUP/high SUB, and 25 with no FI in either muscle. The SUP and SUB were identified in each scan using a machine-learning based automatic segmentation algorithm, after which 93 HU-based radiomic features (18 first-order, 75 texture) were quantified using PyRadiomics[1] (v3.0.1). Additionally, deterministic FI was quantified by dividing the number voxels with HU ranging -190 to -30 (fat) by the number ranging -190 to 50 (fat and muscle). Pearson correlations between Goutallier and each of the image-based measures were performed. Additionally, group-wise comparisons between FI grades (none, low, high) were performed using Wilcoxon rank-sum tests with a Bonferroni correction on the p-value, such that $p < 0.017$ was considered significant.

RESULTS SECTION: The cohort of 100 patients was composed of 49% females with average age and BMI of 70.8±8.6 years and 30.2±6.7, respectively. Degenerative pathologies included 52.0% osteoarthritis, 21.0% rotator cuff tear, and 45.0% cuff tear arthropathy, where multiple diagnoses could be selected. The deterministic FI in the SUP for the none, low, and high groups was 24.1%, 26.1%, and 33.6%, respectively (Table 1). The deterministic FI in the SUB for the none, low, and high groups was 21.2%, 22.7%, and 28.4%, respectively. Correlations of deterministic FI to the Goutallier classification were 0.35 and 0.31 for the SUP and SUB, respectively. Of the 93 HU-based radiomic parameters quantified for each muscle, two had stronger correlations to Goutallier in the SUP, including cluster prominence (0.39) and cluster shade (0.42), and none had stronger correlations to Goutallier in the SUB. Three texture radiomic features had high correlation (range: 0.60-0.62) to deterministic FI in the SUP, while three first-order and 25 texture radiomics had high correlation to deterministic FI in the SUB (range: 0.63-0.80). Regarding pairwise comparisons, there were significant differences in deterministic FI between high vs no FI (p=0.009) and high vs low FI (p=0.004) in the SUP. Six radiomic features were significantly different between high vs no FI, including cluster prominence ($p < 0.001$), cluster shade ($p < 0.001$), cluster tendency (p=0.011), correlation (p=0.013, maximal correlation coefficient (p<0.001), and long run high gray level emphasis (p = 0.009). No features were different between high vs low and low vs no FI. There were no significant differences between groups in deterministic FI in the SUB, however three radiomic features (energy, total energy, and dependence entropy; p=0.001) were significantly different between high vs no FI, three (90^{th} percentile HU [p=0.012], maximum HU [p = 0.003], and HU range [p = 0.011]) were significantly different between high vs low FI, and three (cluster prominence [p=0.004], cluster shade [p=0.012], and small dependence low gray level emphasis [p=0.011] were significantly different between low vs no FI.

DISCUSSION: This pilot study found several radiomic parameters that were more correlated with surgeon-assessed FI than deterministic FI in the SUP, but none in the SUB. Interestingly, the opposite was true considering correlations of radiomic parameters to deterministic FI, where strong correlations were more frequently observed in the SUB than the SUP. Additionally, deterministic FI appeared to distinguish between no, low, and high FI in the SUP, but not in the SUB; this may be a function of the fact that FI is more common clinically in the SUP than the SUB, and therefore more easily subjectively assessed using the Goutallier classification. However, there were several radiomic features that were significantly different between groups in each muscle. Both cluster prominence and cluster shade were more highly correlated to Goutallier than deterministic FI in the SUP. They also distinguished between at least two FI groups in both the SUP and the SUB. Cluster prominence and cluster shade are both features calculated based on the grey level co-occurrence matrix (GLCM), which is a statistical method to examine texture of an image relative to both intensity values and spatial relationships of those intensity values. Cluster prominence is a measure of the skewness and asymmetry of the GLCM, while cluster shade is a measure of skewness and uniformity. Higher values of both of these parameters indicate more asymmetry of the GLCM, potentially capturing the impact of fatty streaking in the muscle.

Though several image-based, automatically quantified radiomic parameters may be potential alternatives to the coarsely discretized, subjective Goutallier classification, the limitations of this pilot study are numerous. First, regarding deterministic FI, HU of similar materials in CT scans may be impacted by image acquisition parameters, such as reconstruction kernel, limiting the ability to define universally accurate HU ranges that correspond to fat and muscle. HU-based radiomic features may be subject to similar limitations. Although multiple reconstruction kernels were included in this study, future work should increase the sample size and simultaneously investigate the effect of image acquisition parameters on potential systematic bias in FI or surrogate values. Second, images used in this study were reconstructed with sharp reconstruction kernels for the purposes of segmenting boney structures for use in pre-operative planning of shoulder arthroplasty. Thus, the soft tissue in these scans may be subject to noise, which is exemplified by >20% deterministic FI on average even in patients that were rated as having no FI using Goutallier. However, these images are also standard of care images, the use of which would be important in any future framework if clinical deployment is a goal.

SIGNIFICANCE/CLINICAL RELEVANCE: A major limitation of current clinical assessment of rotator cuff muscle quality is the subjective nature which necessitates a coarsely discretized scale that is subject to inter- and intra-observer error. Automatic, objective quantification of FI in the rotator cuff muscles may improve treatment-decision making for patients requiring total shoulder arthroplasty, as muscle quality plays a vital role in potential outcomes of different treatment options. Additionally, image-based measures of FI may be applied to other muscles besides the rotator cuff that are not traditionally assessed clinically to determine the impact on clinical outcomes.

REFERENCES: [1] van Griethuysen, J. J. M., et al. (2017). Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Research, 77(21), e104–e107.

Table 1. Significant pairwise comparisons of deterministic FI and radiomic features in the supraspinatus and subscapularis.

	Supraspinatus						**Subscapularis**					
	Group			***p-value***			***Group***			***p-value***		
	High FI	*Low FI*	*No FI*	*High vs Low*	*High vs No*	*Low vs No*	*High FI*	*Low FI*	*No FI*	*High vs Low*	*High vs No*	*Low vs No*
Deterministic FI	33.6	26.1	24.1	**0.009**	**0.004**	0.698	28.4	22.7	21.2	0.084	0.031	0.742
90th Percentile HU	1.09	1.04	1.19	0.320	0.707	0.207	1.02	1.19	1.17	**0.012**	0.146	0.509
Energy	46691	44736	55653	0.707	0.017	0.077	111345	141021	151194	0.146	**0.001**	0.244
Maximum HU	3.90	2.57	3.04	0.134	0.809	0.207	2.53	3.59	3.10	**0.003**	0.114	0.497
HU Range	3.64	2.31	2.90	0.249	0.625	0.187	2.23	3.40	2.91	**0.011**	0.109	0.655
Total Energy	46691	44736	55653	0.707	0.017	0.077	111345	141021	151194	0.146	**0.001**	0.244
Cluster Prominence	5106	4261	2949	0.120	**0.000**	0.068	2074	2219	2109	0.109	0.064	**0.004**
Cluster Shade	141.5	113.4	72.1	0.100	**0.000**	0.057	46.2	53.1	49.0	0.109	0.320	**0.012**
Cluster Tendency	15.7	14.1	14.2	0.128	**0.011**	0.103	11.0	10.4	10.2	0.278	0.070	0.449
Correlation	0.35	0.29	0.26	0.038	**0.013**	0.485	0.26	0.26	0.20	0.617	0.020	0.018
MCC	0.50	0.46	0.43	0.089	**0.015**	0.362	0.41	0.45	0.40	0.259	0.354	0.099
Long Run High Gray Level Emphasis	274.2	237.0	248.6	0.149	**0.009**	0.461	216.5	206.5	237.1	0.422	0.095	0.031
Dependence Entropy	6.30	6.27	6.27	0.100	0.038	0.628	6.17	6.10	6.05	0.036	**0.001**	0.200
Small Dependence Low Gray Level Emphasis	0.002	0.002	0.002	0.259	0.068	0.786	0.002	0.002	0.002	0.131	0.249	**0.011**