Quality assessment and ranking system for quantitative breast diffusion-weighted imaging of the breast in the ACRIN 6698 trial

Sheye Aliu[1], David Newitt[1], Wen Li[1], Jessica Gibbs[1], Lisa Cimino[2], Eunhee Kim[2], Savannah Partridge[3], Patrick Bolan[4], Thomas Chenevert[5], Mark Rosen[6], and Nola Hylton[1]

[1]Radiology & Biomedical Imaging, University of California at San Francisco, San Francisco, CA, United States, [2]ECOG-ACRIN Cancer Research Group, PA, United States, [3]Radiology, University of Washington School of Medicine, Seattle Cancer Care Alliance, WA, United States, [4]Center for Magnetic Resonance Research, University of Minnesota, MN, United States, [5]Radiology, University of Michigan Health System, MI, United States, [6]Radiology, University of Pennsylvania, PA, United States

Purpose: Unique characteristics, such as the varying degree to which adipose tissue is present, and the extent to which poor fat-suppression and artifacts overlap spatially with tissues of interest lead to variable quality in breast diffusion weighted imaging (DWI). The use of strict protocols helps mitigate such variability but does not preclude the need for quality control (QC). In a continuing effort to ensure QC in the ACRIN 6698 trial, an image-grading algorithm was developed by the 6698 Technical Team, a core group of imaging investigators with expertise in breast DWI. It is envisioned that the QC system can be extended for general use in breast DWI and beyond.

Table 1 Criterion used to score the quality of fat suppression, severity of artifacts and SNR

Score	Fat Suppression	Artifacts	SNR
A (4)	Fat is completely suppressed	No artifacts	Contrast along all tissue/background boundaries
B (3)	>75% but <100% of fat is suppressed	1 mild artifact	Contrast along >75% but <100% of boundaries
C (2)	>50% but <75% of fat is suppressed	2 mild artifacts or 1 severe artifact	Contrast along >50% but <75% of boundaries
D (1)	>0% but <50% of fat is suppressed	3 mild artifacts or 2 severe artifacts	Contrast along >25% but <50% of boundaries
F (0)	Total failure of fat suppression	3 severe artifacts	Poor contrast at boundaries
Half scores are used between adjacent rows. Ex: AB indicates a score of 3.5 and captures cases that do not quite fit into A or B.			

Methods: The first step in the QC system is verifying image parameters against specified imaging protocols by reviewing image metadata. Protocol deviations are classified into minor (non-critical to study aims) and major (critical to the study aims). Images exhibiting major protocol violations are omitted from analysis. In the second step, trained readers score images (0-4) on the quality of fat suppression, the severity of artifacts and signal-to-noise ratio (SNR). To allow for granularity, half scores are permitted. See Figure 1 for sample scores and Table 1 for scoring guidance. The geometric mean of the scores from each of the three metrics is computed to yield an intermediate quality grade. A final quality grade is generated by thresholding the intermediate quality grade into high quality, moderate quality or low quality. The third step is a pass/fail assessment of the usability of images for evaluating the diffusion parameters such as the apparent diffusion coefficient (ADC) of lesions. This criterion was established to ensure that images whose quality are adequate to address the scientific aim of the study are not rejected but that their quality is taken into account. Internal testing was conducted to evaluate the utility of the QC system and identify opportunities for improvement. Five readers from the ACRIN 6698 Technical Team participated in testing, were each assigned the same five cases from the 6698 trial to review using the proposed QC system.

Results: For all three metrics evaluated, fat suppression, artifacts and SNR, qualitative comparisons showed good concordance among readers in their assessment of whether images passed or failed for use in evaluating lesions (Figure 2). There was 100% concordance (all 5 readers were in agreement) in 3/5 cases and 80% (4 out of 5 readers were in agreement) concordance in 2/5 cases. Ratings of quality factors were variable among readers. The variance of quality scores was lowest for fat suppression (0, 0.05, 0.18, 0.25, 0.58). Higher variance was observed for artifacts (0.83, 0.38, 0.43, 0.33, 0.43) and SNR (0.5, 1.18, 0.45, 0.08, 0.30).

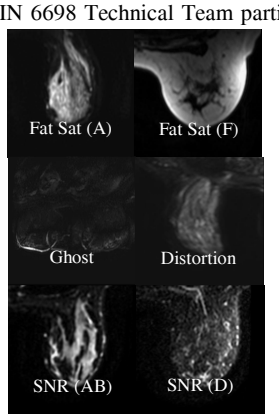

Figure 1 shows representative images that illustrate the scoring for the three QC categories in Table 1. Fat suppression (fat sat) was primarily assessed on T_2 images from the DWI set, i.e. the non-diffusion weighted (b = 0 s/mm^2) images. The top row shows images that were scored A (left) and F (right) for quality of fat-suppression. The severity of artifacts was evaluated using all diffusion-weighted images. Artifacts anticipated to be most debilitating to breast DWI were evaluated: ghosts (middle row, left column), distortion (middle row, right column) and displacement (not shown). SNR was assessed on non-diffusion weighted images. The bottom row shows images that were scored AB (left) and D (right) respectively for SNR.

Discussion: The level of concordance observed when readers assessed the suitability of images for evaluating lesions, serves as preliminary validation of the utility of QC system. While variability was observed among readers in scoring image quality factors, there was very good concordance in determining the pass/fail grade of lesions for measuring ADC value, the quantitative endpoint in the ACRIN 6698 study. The purpose of this study was to perform a preliminary evaluation of the scoring system among a group of expert users. These results will be used for further refinement and testing.

Conclusion: We demonstrate that it is feasible to implement a quality assessment and ranking system for breast DWI, critical to ensuring data quality in multicenter trials. With appropriate improvements, we believe that this QC system is extensible for other applications.

We acknowledge funding support from NCI U01 grants CA079778 and CA080098, and CA151235.

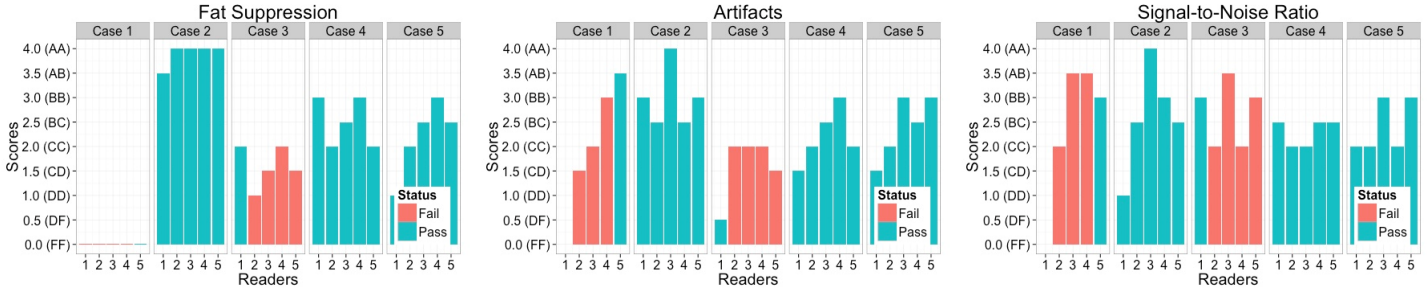

Figure 2 shows results of the scoring of images and the pass/fail assessment of the evaluability of lesions. Each reader is identified on the x-axis (1-5) and his/her quality score (0-4) for each case is shown on the y-axis while his/her pass/fail assessment is color-coded (red = fail, green = pass). These results are shown for fat suppression, artifacts and signal-to-noise ratio. Note that the quality scores for all readers in case 1 for fat suppression was 0, indicating bad fat suppression, and that reader 1 did not provide an artifact or signal-to-noise ratio quality score for case 1.