AbstractID: 5303 Title: Determination of Subjective Similarity for Pairs of Lesions on Mammograms: Comparison of Ranking Scores in 2AFC versus Absolute Ratings for Masses and Microcalcifications

Purpose: We previously obtained the subjective similarity ratings for pairs of lesions on mammograms for quantitative evaluation of similar images. Our purpose in this study was to investigate whether the absolute similarity ratings can be determined reliably by comparison with ranking scores obtained in a 2-alternative forced-choice (2AFC) method.

Method and Materials: We selected 8 pairs of masses and 8 pairs of clustered microcalcifications based on radiologists' average similarity ratings; similarity ratings for the two sets of 8 pairs were approximately evenly distributed. In the first study, each pair was compared one by one to all other 7 pairs in each group of masses and microcalcifications. In the second study, we combined four pairs of masses and four pairs of microcalcifications to compare the similarity of a pair of masses with that of microcalcifications. Seven radiologists and 3 senior residents were asked to choose one pair that was more similar than another pair with the 2AFC method. The cases were presented in randomized order. The number of times that a pair was selected as more similar was counted as the subjective ranking score. The average scores were compared with the average similarity ratings determined previously.

Results: The average ranking scores from the first study were highly correlated (0.93 and 0.98 for masses and calcifications, respectively) with absolute similarity ratings. When mass pairs were compared with calcification pairs, the correlations between ranking scores and absolute ratings were also very high (0.92 and 0.96). In both studies, observers were very consistent in selecting more similar pairs.

Conclusion: The result indicates that absolute similarity ratings determined previously are reliable and useful for selection of similar images. The concept of similarity is robust and meaningful even when mass pairs are compared with microcalcification pairs.