Investigation of Diagnostic and Image Quality Attributes: Comparison of Screen-Film to CR Mammography

Lynn Fletcher-Heath, Anne Richards and Susan Ryan-Kron

Health Group, Eastman Kodak Company
Rochester, NY, U.S.A. 14650-2033

ABSTRACT

Digital mammography is advancing into an arena where analog has long been the gold standard. Direct digital systems may not be the favored solution for a particular site while computed radiography (CR) mammography, remains unproven worldwide. This pilot study responds to the growing desire to acquire and display digital mammographic images by exploring the acceptability of CR mammography. Images representing a range of breast tissue types were collected from 49 subjects (17 screening; 32 diagnostic) at four clinical sites. Comparison views were collected on the same breast, under the same compression, using automatic exposure control on state-of-the-art film systems followed by CR. CR images were processed and printed to a mammography printer for hard copy feature comparison. Each image pair in the study was evaluated according to 13 image quality attributes covering noise, contrast, sharpness, and image quality in the overall captured images as well as in each of several particular breast regions (periphery and skin-line, parenchyma and fatty tissue). A rating scale from 1 to 5 was used (strong preference for film=1, strong preference for CR=5). Twelve experienced mammographers at four clinical sites scored a subset of the 49 cases for a total of 64 image pair readings. There were 64 ratings for each of 13 image quality attributes for all cases and, an additional series of scores (four or five attribute ratings) for each abnormality in the category of mass, architectural distortion and microcalcification, for a total of 1069 scores. Based on the pilot study results, it was suggested that CR was equivalent or preferred to conventional screen-film for overall image quality (38% scored 3; 46% scored >3), image contrast (27% scored 3; 59% scored >3) and sharpness (28% scored 3; 50% scored >3). No preference was found when assessing noise. This pilot study also suggested that diagnostic quality was maintained in assessing abnormalities for attributes necessary to evaluate masses and microcalcifications as compared to screen-film.

Keywords: Screen-film mammography, CR, feature analysis, digital imaging, preference study

1 INTRODUCTION

Though X-ray mammography remains the leading radiographic modality to detect breast cancers at an early stage, the advantages associated with digital modalities have been leading healthcare providers to migrate into an arena where analog has long been the gold standard. There are two major problems associated with this migration: 1) direct digital systems are more expensive than traditional screening clinics can afford and 2) computed radiography mammography systems - though accepted in some parts of the world, remain unproven worldwide. In response to the growing need to acquire and display cost-effective digital mammographic images, this pilot study was conducted, exploring the acceptability of mammography via computed radiography.

1.1 CR Mammography Technology

Computed Radiography has been critically assessed based on intrinsic resolution, noise, and signal-to-noise transfer characteristics and is in worldwide use for general radiography.[1] CR mammography systems have been largely based on general radiography CR systems with various improvements. For example, the Kodak DirectView CR Mammography Feature is an upgrade that enables digital mammography imaging on any one of Kodak's DirectView CR 850, CR 950 or CR 975 systems. These systems and others contain laser scanners capable of reading the latent images formed on a storage phosphor imaging plate and producing a digital image for projection radiography applications.[2] In this case, the mammography feature allows the CR systems to be used for dedicated mammography or mixed-use environments with general radiography.

A new enhanced high-resolution phosphor screen (EHR-M) is provided with the Kodak mammography feature. The new screen has a 56% improved DQE(0) and 34% reduced structure noise compared to the performance of a standard HR (High Resolution) phosphor. The screen is scanned at a pixel pitch of 48.5μm and has an image resolution better than 8 lp/mm. The CR mammography cassettes are available in two standard sizes, 18 x 24 cm and 24 x 30 cm, which is compatible with screen-film mammography x-ray systems.

1.2 CR Mammography Image Processing

Some CR systems have a linear response to incident x-ray exposures and its dynamic range covers up to four decades of exposure levels, greater than that of screen-film mammography systems. The raw capture digital image requires image processing to render it suitable for optimal display. Image processing can compress the image dynamic range, selectively enhance image features of various sizes, and apply a tone-scale to achieve proper brightness and contrast in the final rendered image. Figure 1 demonstrates the difference in appearance between CR and screen-film mammography via the same view and compression of a CC view. CR overcomes the traditional analog film trade-off between exposure and detail contrast[3], the full exposure range of anatomy within digital mammography images can be viewed optimally without loss of detail contrast.

Commonality exists among most image processing algorithms necessary to display digital radiographs. Some vendors first convert their data into a logarithmic space such that a linear response with the log of incident exposure is represented. The minimum processing step is to create and apply a tone scale, which restores a film-like appearance to the linear or log exposure data. The next step is to accentuate interesting frequency ranges. Although implementations vary, details are usually emphasized while dynamic range is compressed. Flexibility in the implementation to accommodate various anatomies and body positions may also be incorporated.

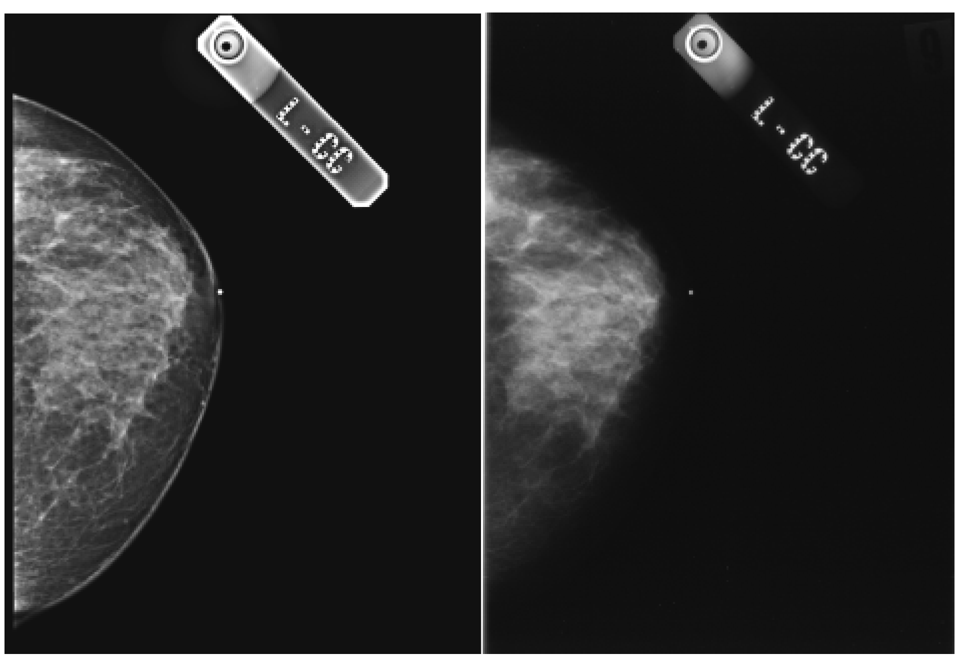

Figure 1: (left) Computed radiography image of the breast, (right) Conventional analog image of the same breast and compression as on the left.

2 MATERIALS AND METHODS: THE CLINICAL PROTOCOL

2.1 Image Acquisition

All images were acquired from one of four clinical sites following a study protocol designed in accordance with Good Clinical Practices for Medical Devices. Images were collected with informed consent from 49 subjects representing a mixture of screening and diagnostic cases and a range of breast tissue types. The diagnostic patient populations consisted of 32 cases with 16 masses, 18 groupings of microcalcifications (MCCs) and 7 architectural distortions.

All subjects had the standard craniocaudal (CC) and mediolateral oblique (MLO) views in screen-film mammography as it was typically performed at the facility. The default screen-film systems at these clinical sites were one of Kodak Min-R L, Min-R 2000 or Min-R EV. Standard techniques for each site were used based on the calibration of the on-site automatic exposure control and the screen-film system in use. A controlled experiment measured exposure differences between screen-film and CR. The use of AEC and Lucite (2, 4, 6, 7 cm) simulating various breast compressions yielded an average 15% less exposure using CR than the Min-R EV film system developed in EXII chemistry.

Digital CR mammography images were acquired on the Kodak DirectView CR 850 system with the Mammography Feature enabled. Two views (CC, MLO) were taken for one breast for each subject. To minimize variations each CR image was taken immediately following the corresponding screen-film image while the breast was still compressed. Images were processed using site approved defaults for Kodak PTS, and EVP software then printed using either the Kodak DryView 8610 or the Kodak DryView 8900 Mammography Laser Imager and DVM film. The printed film density range was from 0.21 to 3.5 for the DryView 8610 and 0.21 to 3.6 for the DryView 8900.

The printed CR images were used in this study as test images, and their corresponding screen-film images were considered controls. All observer data for this study was obtained given a comparative feature analysis of hard copy matching views – CR to screen-film for each of the two views for each subject.

2.2 Clinical Evaluation

Twelve radiologists from the four clinical sites participated in the study. All have had extensive experience in mammography with varying levels of digital experience. Each radiologist reviewed a subset of the 49 subjects in a paired presentation of the screen-film and CR images (same breast, same view from the same patient) with the viewing conditions in use at the facility (dedicated mammography view boxes, masking, and low ambient lighting) resulting in a total of 64 image-pair readings. We define one reading as all data collected from one radiologist viewing one image pair. A case read by more than one radiologist was considered as an additional independent read for this reporting. The film image was placed on the left side and the CR image on the right side. Left/Right randomization was not done, as the images were easy to differentiate. A score was provided in a total of 13 categories for each pairing according to their impressions of overall noise, contrast, sharpness and image quality and for the noise contrast and sharpness in each of the breast regions (periphery and skin-line, parenchyma and fatty tissue). A rating scale from 1 to 5 was used, with a score of 1 corresponding to a definite preference for the screen-film image and a score of 5 indicating a definite preference for the CR image. A score of 3 indicated the CR image was equivalent to the conventional analog radiographic film (Table 1). There were 64 scores for each of 13 image quality attributes (832 scores) plus four attributes were scored for each architectural distortion and five attributes were scored for each mass and each microcalcification. The suspicious findings were comprised of 16 masses, 18 MCCs and 7 architectural distortions (total 41). Second readings from additional radiologists enhanced the number to 18 reads of masses, 23 reads of MCCs and 8 reads of architectural distortions (total 49). Scores for suspicious findings totaled 237, categorizing: margin sharpness, lesion contrast, density and ability to diagnose masses; shape, count, sharpness, distribution and ability to diagnose MCCs; density, parenchymal edge and ability to diagnose architectural distortions.

Table 1: The feature analysis rating scale

RATING	EXPLANATION
1	Film image is diagnostically superior
2	Film image is better
3	No difference between film image and CR image
4	CR image is better
5	CR image is diagnostically superior

Each radiologist also assessed the images according to the BI-RADS lexicon for breast parenchyma density as: almost entirely fatty; scattered fibroglandular density; heterogeneously dense; and, homogeneously dense. All abnormalities categorized as MCCs, masses and architectural distortions were reviewed and scored using the same feature analysis rating scale.

3 RESULTS

3.1 Diagnostic ratings based on image quality preference

Statistical treatment of the observer data demonstrates a few interesting points that will be correlated with trend analysis.[4,5] Table 2 shows the rated attributes separated into four major categories: (1) the overall image, (2) the breast periphery and skin line, (3) the breast parenchyma and (4) the fatty tissue. Note the mean ratings are all above the center value of 3 (indicating preference towards the CR image), as are all lower confidence limits based on a 95% confidence limit. Also, as one might expect, higher ratings existed for the general category of the breast periphery and skin line, as this area is better visible on the CR image due to dynamic range compression processing. Also note that the lowest valued attribute, though still above 3, is the noise level within each general category.

Table 2: Observer data for 13 image quality attributes. Bolded are the Mean and Lower Confidence Limit based on a 95% confidence level. Recall values above 3 indicate a preference towards the CR image.

	Overall Image Attributes				**Breast Periphery/Skinline**			**Breast Parenchyma**			**Fatty Tissue**		
	Overall Image Quality	*Res/ Sharp*	*Overall Contrast*	*Noise Levels*	*Res/ Sharp*	*Contrast*	*Noise*	*Res/ Sharp*	*Contrast*	*Noise*	*Res/ Sharp*	*Contrast*	*Noise*
Mean	**3.64**	**3.61**	**3.78**	**3.36**	**4.21**	**4.31**	**3.74**	**3.55**	**3.71**	**3.52**	**3.78**	**3.89**	**3.55**
Median	3.75	4.00	4.00	3.00	4.00	4.00	4.00	4.00	4.00	3.00	4.00	4.00	3.00
Mode	3.00	4.00	4.00	3.00	4.00	4.00	3.00	3.00	4.00	3.00	4.00	4.00	3.00
Standard Deviation	0.87	0.99	0.81	1.00	0.72	0.58	0.86	0.98	0.92	0.87	0.87	0.78	0.82
Confidence Level(95.0%)	0.22	0.24	0.20	0.24	0.17	0.14	0.21	0.24	0.22	0.21	0.21	0.19	0.20
Lower Confidence Limit	**3.42**	**3.36**	**3.58**	**3.12**	**4.03**	**4.17**	**3.53**	**3.31**	**3.49**	**3.31**	**3.56**	**3.70**	**3.34**

Since the **Overall Image Attributes** category represents the general view for the entire breast, these are the ratings that will be discussed in the remainder of this section and shown in Figures 2 – 6.

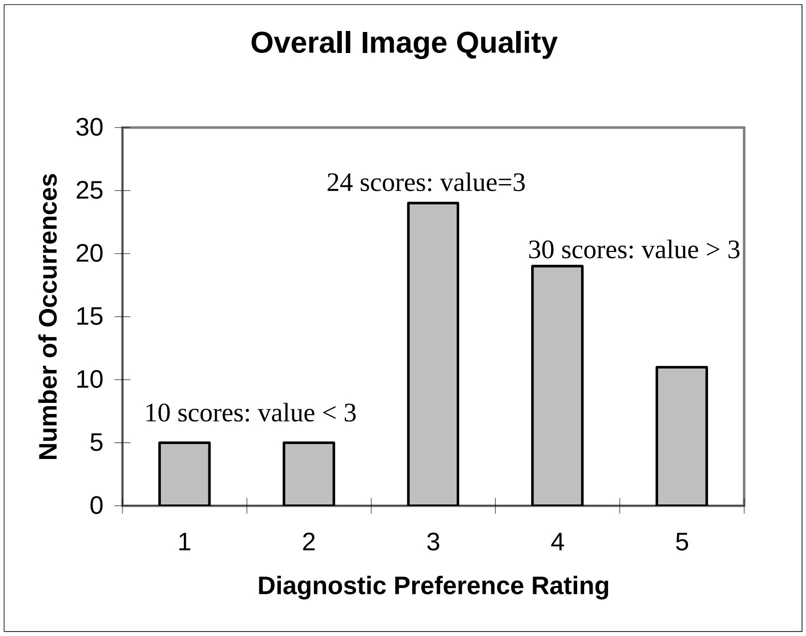

Figure 2: Distribution of 64 diagnostic preference ratings based on overall image quality for 49 cases.

The overall image quality distribution in Figure 2 is based on ratings specific to the entire breast region of all 64 reads of the 49 cases. Each radiologist indicated one diagnostic preference rating for each film/printed CR pair observed. A chart balanced over the rating 3 would indicate no preference to either modality, however, Figure 2 demonstrates a significant skew towards the CR image appearance. Screen-film was preferred 10 occurrences of 64 or 16% of the time, CR was preferred 24 occurrences of 64 or 46% of the time and no preference was noted on 24 occurrences or 38% of the time.

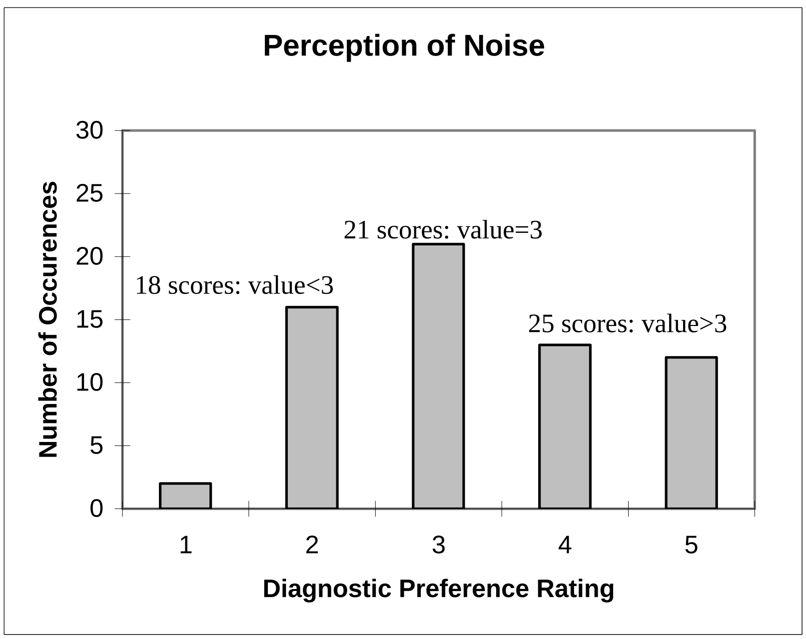

Figure 3: Distribution of 64 diagnostic preference ratings based on the appearance of noise for 49 cases.

Figure 3 demonstrates a slight, though statistically insignificant preference for the CR image with respect to noise. Screen-film was preferred 18 of 64 occurrences or 28% of the time, CR was preferred 25 of 64 occurrences or 39% of the time and no preference was noted 21 occurrences of 64 or 33% of the time. Figure 4 indicates a preference towards the CR image for the attribute of sharpness. This may be due to the spatial frequency enhancement processing, which is unavailable in screen-film systems. Screen-film was preferred 14 of 64 occurrences or 22% of the time, CR was preferred 32 of 64 occurrences or 50% of the time and no preference was noted 18 of 64 occurrences or 28% of the time.

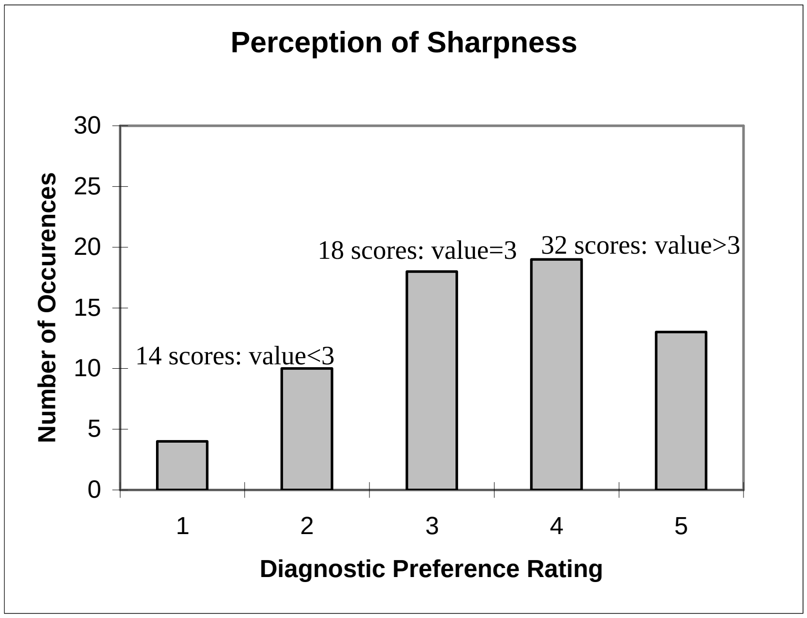

Figure 4: Distribution of 64 diagnostic preference ratings based on perceived sharpness for 49 cases.

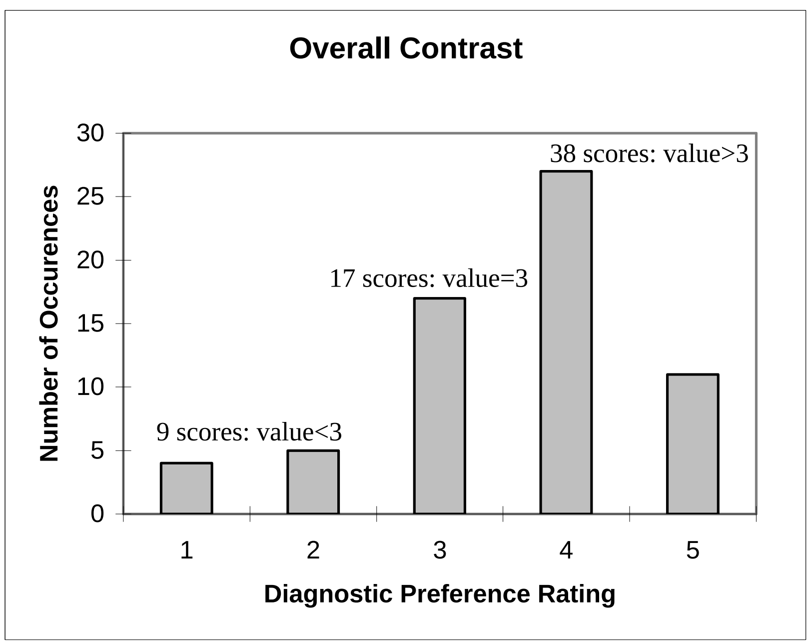

Figure 5: Distribution of 64 diagnostic preference ratings based on overall contrast for 49 cases.

Another notable result is displayed in Figure 5, where an overall preference existed for the CR image with respect to the overall contrast; again, this may be directly attributable to the availability of image processing. In this case, screen-film was preferred 9 of 64 occurrences or 14% of the time, CR was preferred 38 of 64 occurrences or 59% of the time and no preference was noted 17 of 64 occurrences or 27% of the time.

Breast type had an impact on the degree of preference as exhibited by Figure 6. Though benefit existed in the fatty to heterogeneously dense categories (55%, 52%; 59% respectively), a stronger preference existed for the homogeneously dense category (71%) for Overall Image Quality. A common criticism of screen-film mammography is that penetration of dense glandular tissue is difficult.[6] Digital image processing offers a solution to this problem by compressing the dynamic range of the image while enhancing the details necessary to make a diagnosis. This may not only impact the ability to diagnose within dense breast tissue but may also reduce the number of exposure retakes.

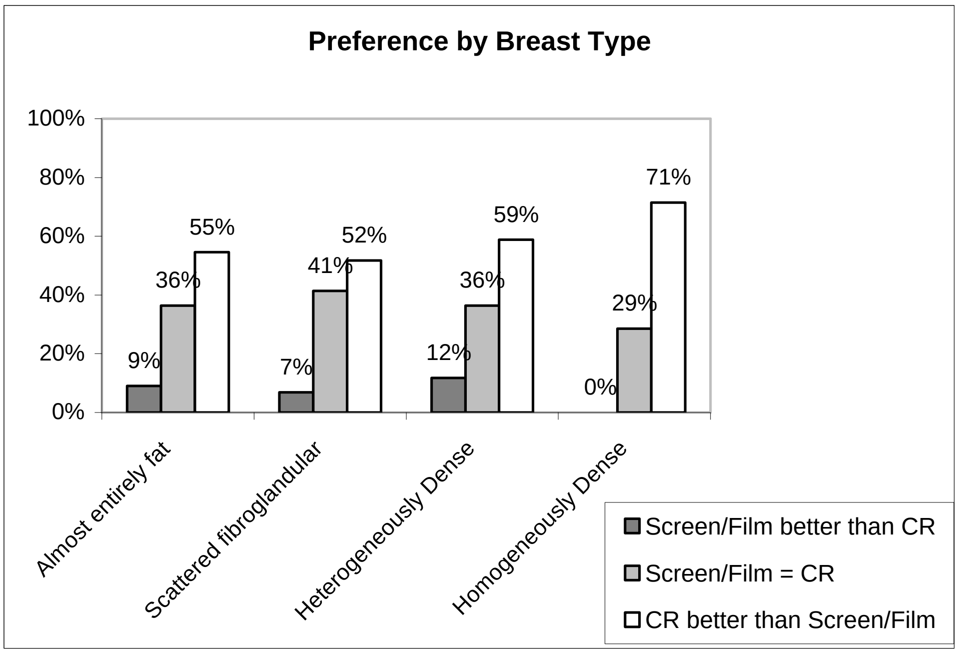

Figure 6: Percent of each image pair separated by BI-RADS tissue categorization based on preference ratings less than 3 (Screen-film preferred), equal to 3 (equal preference) and greater than 3 (CR preferred) for Overall Image Quality.

3.2 Diagnostic Ratings of Suspicious Findings

A further separation of only the diagnostic patients presents another important aspect of required diagnostic quality, shown in the next set of figures. Forty-nine series of additional ratings were captured for each of 16 masses, 7 architectural distortions and 18 groupings of MCCs. Each suspicious finding was subjected to a relevant series of questions corresponding to the characteristics of each type of finding. Though more data would be required to support full statistical scrutiny, trends may be observed for each attribute, for each type of finding, for each modality. Figures 8 – 10 demonstrate all sub-categorical attributes of the finding type while Figure 10 shows a comparison of the assessment "Ability to make a diagnosis" based on the appearance of all findings.

Distribution of mass attributes for the 18 readings is shown in Figure 7. The attributes assessed were shape, margin sharpness, lesion contrast and density. The clearest comparison was for the attribute of shape. A "no preference" rating was given 14 of 18 times, screen-film and CR are preferred 2 times each. Margin sharpness is preferred on the screen-film image 7 times, and preferred on CR 4 times. No preference existed 7 times. Similarly, lesion contrast (S/F 6, CR 2; no preference 10) and density ratings (S/F 6, CR 2; no preference 10) were also shifted towards the screen-film rendering, albeit slightly. The margin sharpness, lesion contrast and density of masses captured here tend to be preferred on screen-film. Since optimization of image processing was not done during the study, or at all based on abnormal findings, further study is required.

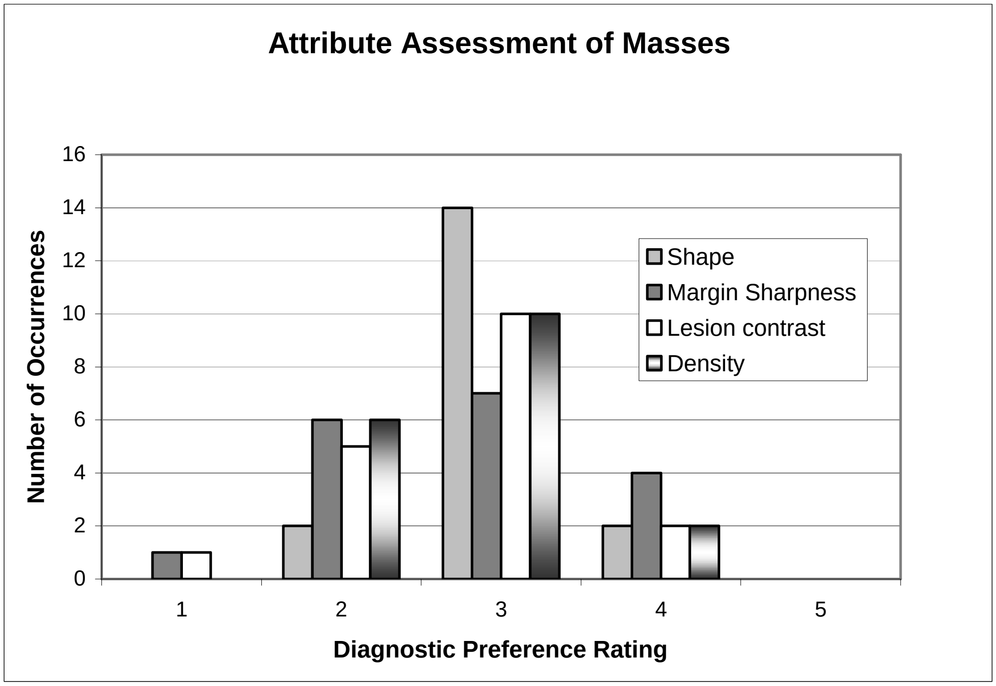

Figure 7: Distribution of the available 18 ratings of mass attributes for 16 masses. Preference to screen -film is indicated by preference ratings less than 3, no preference is indicated by ratings equal to 3 and Preference to CR is indicated by ratings greater than 3.

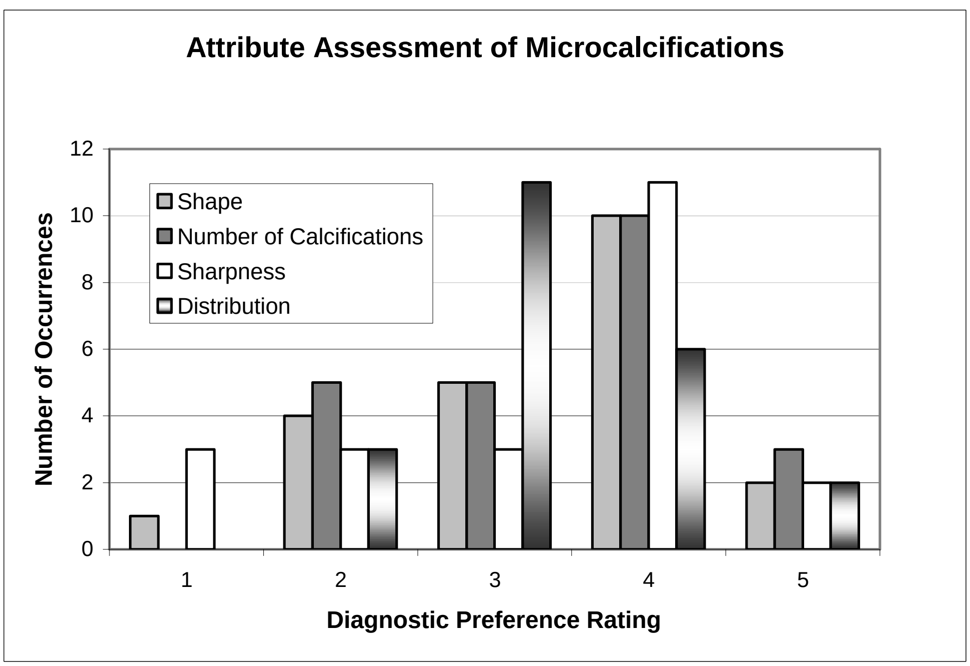

Figure 8: Distribution of the available 23 ratings of microcalcification attributes for 18 groupings of microcalcifications.

Distribution of microcalcification attributes for the 23 readings is shown in Figure 8. The attributes assessed were shape, count, sharpness, and distribution. Preferences towards the CR rendering are clear for all attributes by which MCCs are assessed. Shape, count, and sharpness (median score is 4) are more clearly weighted towards CR than the distribution attribute (median score is 3); however, this attribute yields 11 ratings with "no preference," 8 preferred with CR and 3 with preference towards screen-film. Examining the tails of the chart, 2 MCCs rated "5" and none rated "1."

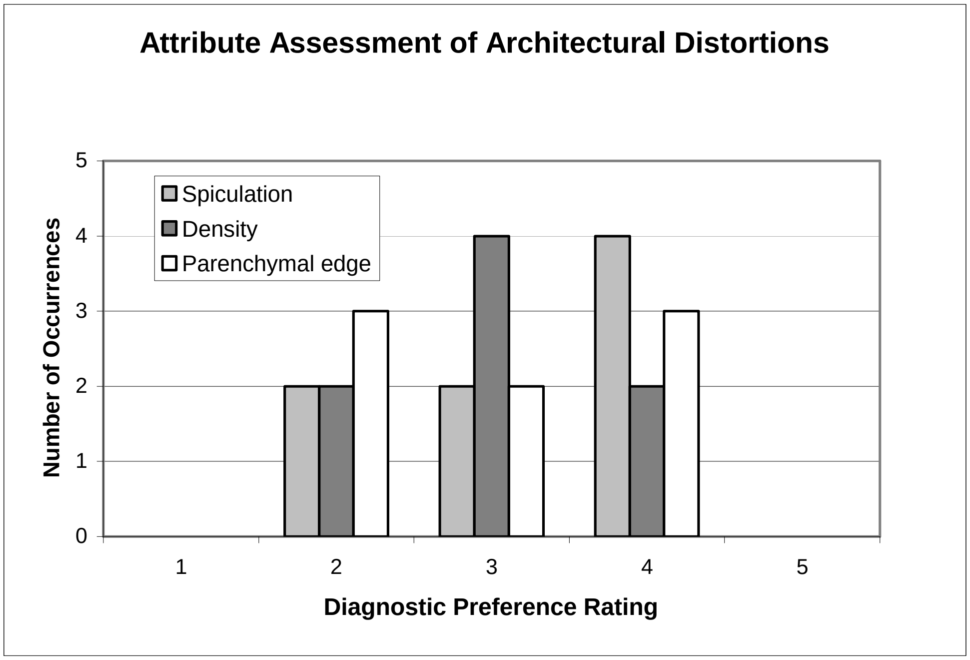

Figure 9: Distribution of the available 8 ratings of attributes for seven architectural distortions.

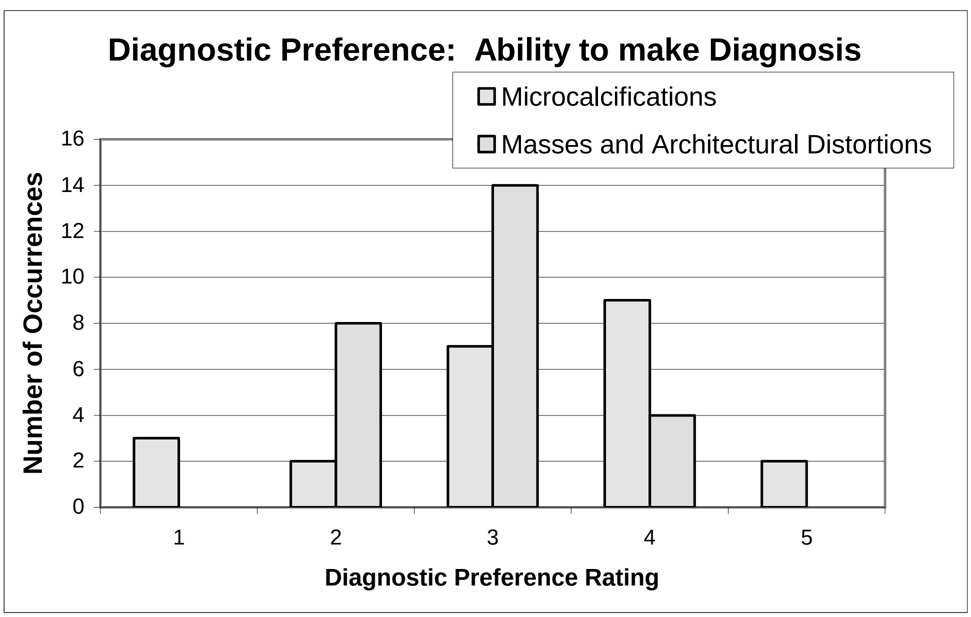

Figure 10: Modality preference based on observations of masses, architectural distortions and microcalcifications. Due to the small number of architectural distortions, they were combined with masses.

The ratings obtained regarding architectural distortions were in assessment of spiculation, density and parenchymal edge. Due to the small number of architectural distortions, the most that should be stated is that the combination of these attributes yields a chart nicely centered on the preference rating of 3, or "no preference." The final representation of all diagnostic data (Figure 10) combines masses and architectural distortions due to the amount of available data.

Figure 10 shows that a preference towards CR was observed when assessing MCCs and a slight preference towards the screen-film was observed when assessing masses and architectural distortions. While assessing MCCs, CR was preferred 48% of the time, screen-film was preferred 22% of the time and no preference was noted 30% of the time. In the assessment of masses and architectural distortions, CR was preferred 15% of the time, screen-film was preferred 31% of the time and no preference was noted near 54% of the time. In other words, radiologists perceived diagnostic quality was maintained or improved 78% for MCC and maintained or improved 69% for masses while of those that were preferential towards the screen-film rendering, only three ratings of the 49 scored a "1" equating to "screen-film being superior."

4 DISCUSSION AND CONCLUSIONS

A pilot clinical study has been described that suggests evidence of a diagnostic look preference for the digital CR mammography images, over screen-film images of the same breast compression for the attributes of overall image quality, perceived sharpness/resolution and contrast. No preference was found for the attribute of noise for either modality. Separating preference by breast type indicated that there is an increasing benefit provided by the CR system as the breast tissue increases in density. This finding is correlative to the results of the Digital vs. Film Mammography in the Digital Mammographic Screening Trial (DMIST).[6]

The data from the diagnostic cases, though limited, demonstrates the ability of CR mammography to present image features preferable to state-of-the-art screen-film systems. This capability, especially given the use of default parameters, which were untuned to mammographer preference or the appearance of abnormalities, shows promise and exploration of preferences will continue.

Some study limitations require discussion. There were a limited number of cases, which resulted in a limited number of readings. Since, most screen-film images were required to stay in the acquiring clinic, few pairs could be presented to mammographers in other clinics. Also, there is a possibility that the available repeat readings may bias results since the repeat readings were not generated for all cases. In addition, this design offers no immediate access to analyze inter and intra-observer variability since each reading was treated as an independent occurrence.

Each site used the screen-film system that was already in use at that site to maximize mammographers' comfort with the comparison. A limitation of this is the increased number of variables in the study where one of three possible screen-film systems was compared to the CR acquisition. It would seem optimal to reduce as many variables as possible, especially when data may be limited.

Image processing plays a significant role in the presentation of optimal renderings. The on-going discussion about required acceptance and verification tests for digital mammography systems is warranted. Judging a system by its final output can be done easily for screen-film imaging systems where final assessment is done on the developed film; for digital systems the path is not as clear, as image processing algorithms are generally designed to work optimally for anatomy, not phantoms. Ultimately, studies with expert mammographers will provide statistics of diagnostic accuracy as the true measure of system performance. Results from this pilot study do show various trends in favor of the use of CR for mammography, and though diagnostic preference does not guarantee improved diagnostic accuracy, it can improve diagnostic confidence. Further studies can more accurately address these concerns.

5 ACKNOWLEDGMENTS

The authors wish to sincerely thank the technologists, physicists and mammographers of the following clinics for their support during this study - MeritCare Health System (Fargo, ND, U.S.A.), University of Rochester Medical Center (Rochester, NY, U.S.A.), Clinica Santa Maria (Santiago Chile), and Tomovale Centro de Diagnostico por Imagem (Sao Jose dos Campos, Sao Paulo, Brazil). We would especially like to thank Brent Colby for his dedication to quality assurance and for his dose characterization efforts. We would also like to express our deepest gratitude for the opportunity to work with Dr. Jeanne Cullinan.

References:

[1] E. Samei and M. Flynn. Physical Measures of Image Quality in Photostimulable Phosphor Radiographic Systems. *Proc SPIE*, 1997; 3032:328-338.

[2] T. Bogucki, D. Trauernicht and T. Kocher. *Characteristics of a Storage Phosphor System for Medical Imaging. Health Sciences Division*, Eastman Kodak Company, Rochester, NY 14650, 1995. Technical and Scientific Monograph No. 6.

[3] R. Van Metter and D. Foos. Enhanced latitude for digital projection radiography. *Proc. SPIE*, 1999; 3658:468-483.

[4] D. Clason and T. Dormody. Analyzing data measure by individual Likert-Type items. Journal of Agriculture Education, 1994; 4:31-35.

[5] Jaccard, James and Choi K. Wan (1996). *LISREL approaches to interaction effects in multiple regression.* Thousand Oaks, CA: Sage Publications.

[6] E. Pisano *et.al.* Diagnostic Performance of Digital versus Film Mammography for Breast-Cancer Screening. *The New England Journal of Medicine*, 2005; 353: 1773-1783.