Radiologists' Preferences for Digital Mammographic Display[1]

Etta D. Pisano, MD
Elodia B. Cole, MS
Stacey Major, MS
Shuquan Zong, MS
Bradley M. Hemminger, MS
Keith E. Muller, PhD
R. Eugene Johnston, PhD
Ruth Walsh, MD
Emily Conant, MD
Laurie L. Fajardo, MD
Stephen A. Feig, MD
Robert M. Nishikawa, PhD
Martin J. Yaffe, PhD
Mark B. Williams, PhD
Stephen R. Aylward, PhD
For the International Digital Mammography Development Group

Index terms:
Breast, abnormalities, 00.30, 00.81
Breast radiography, comparative studies, 00.112, 00.1215, 00.1299
Breast radiography, technology, 00.1215, 00.1299
Cancer screening
Images, display
Images, processing

Radiology 2000; 216:820–830

Abbreviations:
ANOVA = analysis of variance
CLAHE = contrast-limited adaptive histogram equalization
HIW = histogram-based intensity windowing
MIW = manual intensity windowing
MMIW = mixture model intensity windowing
MUSICA = multiscale image contrast amplification

[1] From the Dept of Radiology, University of North Carolina CB7510, 503 Old Infirmary, Chapel Hill, NC 27599-7510 (E.D.P.). Affiliations for all other authors and members of the International Digital and Mammography Development Group are listed at the end of this article.
Received Oct 5, 1999; revision requested Nov 5; revision received Dec 17; accepted Jan 28, 2000. Financial support is listed at the end of this article.
Address correspondence to E.D.P. (e-mail: *Etta_Pisano@med.unc.edu*).

Author contributions:
Guarantor of integrity of entire study, E.D.P. A complete list of author contributions appears at the end of this article.

PURPOSE: To determine the preferences of radiologists among eight different image processing algorithms applied to digital mammograms obtained for screening and diagnostic imaging tasks.

MATERIALS AND METHODS: Twenty-eight images representing histologically proved masses or calcifications were obtained by using three clinically available digital mammographic units. Images were processed and printed on film by using manual intensity windowing, histogram-based intensity windowing, mixture model intensity windowing, peripheral equalization, multiscale image contrast amplification (MUSICA), contrast-limited adaptive histogram equalization, Trex processing, and unsharp masking. Twelve radiologists compared the processed digital images with screen-film mammograms obtained in the same patient for breast cancer screening and breast lesion diagnosis.

RESULTS: For the screening task, screen-film mammograms were preferred to all digital presentations, but the acceptability of images processed with Trex and MUSICA algorithms were not significantly different. All printed digital images were preferred to screen-film radiographs in the diagnosis of masses; mammograms processed with unsharp masking were significantly preferred. For the diagnosis of calcifications, no processed digital mammogram was preferred to screen-film mammograms.

CONCLUSION: When digital mammograms were preferred to screen-film mammograms, radiologists selected different digital processing algorithms for each of three mammographic reading tasks and for different lesion types. Soft-copy display will eventually allow radiologists to select among these options more easily.

Digital mammograms can be printed on film or displayed on a monitor. Typically, laser-printed films can display 4,000 × 5,000 pixels at 12-bit gray scale. Although most radiologists are currently more comfortable with the use of these printed images, the disadvantages of hard-copy image display at digital mammography are obvious. Once an image is printed, it can no longer be manipulated, and any information available in the digital data but not captured on the printed image will therefore be lost.

With the currently available high-luminance high-spatial-resolution monitors (2,000 × 2,500 pixels) (1), only a portion of the breast can be displayed at one time with full resolution on many digital mammograms. In addition, comparison of prior images with current images and of left images with right images is difficult. Use of roaming, zooming, and gray-level manipulation of the digital images with the computer, while possible, is not trivial to learn and can be inefficient and time-consuming. Memory requirements for online interpretation are currently prohibitive. More practical displays with short, clinically acceptable display times for the entire set of images, including comparison images, are needed before digital mammography can reach its full potential. Exploration of this issue was the purpose of a recent meeting of a working group jointly sponsored by the Office of Women's Health, Washington, DC, and the National Cancer Institute, Bethesda, Md (2).

Given the preferences of radiologists and the present limitations of soft-copy technology, digital mammograms will most likely be displayed on film for the next few years at

least. Therefore, exactly how the images should be printed is an important issue. Even if soft-copy display is used, it is important to determine how the images should be viewed for optimal visualization of different lesion types in breasts of different radiographic densities.

To the best of our knowledge, this is the first study in which the utility of displaying printed digital mammograms by use of eight different image processing algorithms was systematically explored. We sought to determine the preferences of radiologists for algorithms used in the two main tasks in mammography: lesion detection (screening) and characterization (diagnosis).

MATERIALS AND METHODS

Image Production

Radiologist investigators (E.D.P., L.L.F., D.B.K., and E.C.) at four participating institutions selected unilateral digital mammograms for inclusion in the study. Images were deemed eligible for inclusion if mammographic findings were present and if the screen-film image obtained in the same patient was available for comparison. The cases were obtained by using three different full-field digital mammographic devices: 10 cases were obtained with the Trex Digital Mammography System (Trex Medical, Long Island, NY); 10 cases, the SenoScan (Fischer Imaging, Denver, Colo); and eight cases, the Senographe 2000D (GE Medical Systems, Milwaukee, Wis).

Study cases were selected from all digital mammograms available at the involved institutions. Co-investigators were asked to select mammograms that contained findings (ie, masses or calcifications), preferably in patients with dense breasts. The findings were either histologically proved or were considered benign by virtue of mammographic stability for at least 1 year. The goal was to obtain 10 cases for each manufacturer, but only eight cases were obtained with the GE Medical Systems unit.

The raw digital data were transmitted to the University of North Carolina, Chapel Hill, and to other participating institutions for image processing purposes by means of 8-mm tape (Exabyte, Boulder, Colo) or the Internet by using file transfer protocols. Tapes were read by using an 8-mm tape drive (Exabyte).

Images obtained with the Trex system were 4,800 × 6,400 pixels, with a 40-μm pixel. Images obtained with the GE Medical Systems unit were 1,800 × 2,304 pixels, with a 100-μm pixel. Images obtained with the Fischer Imaging unit were 3,072 × 4,800 pixels, with a 50-μm pixel. All three units produced images with 16 bits per pixel.

All images were processed by using each of eight different algorithms: manual intensity windowing (MIW), histogram-based intensity windowing (HIW), mixture model intensity windowing (MMIW), contrast-limited adaptive histogram equalization (CLAHE), multiscale image contrast amplification (MUSICA; Agfa Division of Bayer, Ridgefield, NJ), unsharp masking, peripheral equalization, and Trex processing. Details regarding how these algorithms were applied in this study are described in the Appendix.

During processing, the original contrast and spatial resolution were maintained on all images. Images processed with HIW, MIW, MMIW, and Trex algorithms were printed on film without subsequent contrast manipulation of any type. Before printing, an experienced mammographic technologist manually set the intensity windows on images processed with CLAHE, peripheral equalization, and unsharp masking. Intensity windows on MUSICA-processed images were set over a fixed range (0–4,095 gray values). A single high-brightness (100 footlamberts) monitor (model 1654; Orwin Associates, Amityville, NY) with an Md5 Sun Display Card (Dome Imaging, Waltham, Mass) and an UltraSparc computer (model 2200; Sun Microsystems, San Jose, Calif) were used for all manual intensity windowing. Both the monitor and display card had a display matrix size of 2,048 × 2,560 pixels.

All images except those that underwent Trex processing were printed on film (Ektascan HN; Eastman Kodak, Rochester, NY) by using a laser film printer (K2180 Ekta-Scan; Eastman Kodak). This printer was capable of printing 12 bits per pixel. Images that contained a bit range wider than that of the printer were linearly remapped to the range of the printer. Images were bilinearly interpolated by the printer to its maximum spatial resolution, with a 50-μm pixel size and a matrix of 4,096 × 5,120, and were printed by using the printer at this resolution. The laser film was processed by using a medical film processor (QX-400; Konica Medical, Norcross, Ga).

Images processed with the Trex algorithm were printed on helium-neon film (Scopix LT-2B; Agfa) by using a film printer (LR5200; Agfa). This printer was capable of printing 8 bits per pixel. The matrix size for was 4,776 × 5,944 pixels, with a 40-μm pixel size. Films were processed by using a processor (RP-Xomat; Eastman Kodak).

Mammograms obtained with the Trex system were cropped from 4,800 × 6,400 pixels to fit the printer matrix size. Images obtained with the GE Medical Systems and Fischer Imaging units were scaled up by using interpolation with factors of 1.35 and 3.50, respectively. All printers and monitors used in this study were calibrated to comply with the Digital Imaging and Communications in Medicine gray-scale display function standard (3).

Of the 224 processed digital mammograms, three (1.3%) could not be printed because of printer software errors. These were excluded from the study. In addition, an experienced breast imager (E.D.P.) who did not participate in the ranking of the images reviewed the quality of all printed digital images. Four (1.8%) images were deemed to have very poor quality and were excluded for that reason. Thus, seven (3.1%) of the original 224 digital mammograms prepared were not included in the study.

Preference Study

A total of 65 lesions were identified and were circled on the two views of a single version of the digitally printed image of the patient's digital mammogram. A written description of each of the circled lesions was prepared. This description included histologic information about the lesion if it was available. Other lesions were presumed to be benign by virtue of a minimum of 1 year of mammographic stability with no clinical findings. Tables 1–3 list the images included in this study. Figures 1–4 show representative cases. Each rated case had one to six lesions to evaluate. Cases included only pathologically proved lesions (two obtained with the GE Medical Systems unit; five, Trex unit; and two, Fischer Imaging unit), only presumed benign lesions (three obtained with the GE Medical Systems unit; and five, Fischer Imaging unit), or both types of lesions (three obtained with the GE Medical Systems unit; five, Trex unit; and three, Fischer Imaging unit).

Twelve radiologists (including M.P.B., R.M., R.W., E.R., M.S.S., M.W., A.V., P.J.K.), all qualified mammographic interpreters according to the Mammography Quality Standards Act, independently participated as readers in this study. Written instructions were provided to each radiologist prior to the study. Appropriate masking of the viewboxes was used throughout.

TABLE 1
Case Types: Pathologically Proved Lesions (*N* = 29)

Machine Manufacturer	No. of Cases	No. of Masses		No. of Calcifications		No. of Architectural Distortions	
		Cancerous	Noncancerous	Cancerous	Noncancerous	Cancerous	Noncancerous
GE Medical Systems	5	4	2	1	1	0	0
Trex Medical	10	4	5	2	4	0	0
Fischer Imaging	5	1	2	1	1	0	1

A research assistant (E.B.C.) presented the 28 cases to each reader in random order. The craniocaudal images of each patient were presented first, followed by the mediolateral oblique images. The eight processed digital mammograms were randomly presented in each case. Readers were also provided with the corresponding screen-film mammogram in the same patient, the annotated printed digital mammogram of the same view (lesions circled and numbered), and the description of the histologic findings in each case. The radiologists hung the annotated image on the top viewbox panel of a standard mammographic lightbox (two-tier desktop Mammography Illuminator; Picker, Norcross, Ga). The screen-film mammogram and one of the eight processed digital mammograms to be rated were hung on the lower viewbox panel. Radiologists were provided with a magnifying glass and were encouraged to use it.

First, radiologists were asked to rate the visibility and characterizability of each lesion on the processed digital image with respect to its depiction on the corresponding screen-film mammogram. Radiologists were instructed to use their experienced judgement in determining which areas on the screen-film image corresponded to the lesions depicted on the digital images, while taking into account differences in positioning, compression, and other factors. Using all relevant clinical information, readers were asked to consider whether the processed digital image allowed sufficient visualization and characterization of each lesion so that the correct diagnosis could be reached. Each lesion on the digital mammogram was rated on a five-point scale as much better, better, the same, worse, or much worse than its screen-film counterpart (+2, +1, 0, −1, or −2, respectively) with respect to visibility and characterizability. No magnification images or spot radiographs were provided to the readers.

Next, the radiologists were asked to rate the processed digital image as much better, better, the same, worse, or much worse than the corresponding screen-film image for the purpose of screening (+2, +1, 0, −1, or −2, respectively). For this task, they were asked to consider whether the digital image allowed sufficient visualization of all relevant anatomic structures for effective breast cancer screening. They were instructed to disregard artifacts that occurred outside the borders of the breast in making this judgment. Again, craniocaudal images were rated first, followed by mediolateral oblique images.

The radiologists completed the tasks in the order in which they were presented. To limit the effects of fatigue, short breaks (at least 5 minutes) were required after every 50 minutes of work. The radiologists also took additional breaks as needed. On average, the radiologists required 5 hours to evaluate all images.

TABLE 2
Case Types: Lesions Presumed to Be Benign Due to Stability (*N* = 36)

Machine Manufacturer	No. of Cases	No. of Masses	No. of Calcifications	No. of Architectural Distortions
GE Medical Systems	6	7	5	1
Trex Medical	5	3	3	1
Fischer Imaging	8	5	10	1

TABLE 3
Case Types: Histologic Diagnoses for Pathologically Proved Lesions

Diagnosis	No. of Masses	No. of Calcifications	No. of Architectural Distortions
Cancers			
Infiltrating ductal carcinoma	6	0	0
Ductal carcinoma in situ	1	4	0
Infiltrating lobular carcinoma	2	0	0
Noncancers			
Atypical ductal hyperplasia	1	2	0
Fibrocystic change	1	2	0
Cyst	3	0	0
Fibroadenoma	2	0	0
Lobular carcinoma in situ and atypical lobular hyperplasia	1	1	0
Intraductal papilloma	1	0	0
Atrophy	0	1	0
Chronic inflammation and fibrosis	0	0	1

Note.—The total number of pathologically proved lesions is not equal to the number of cases because some cases had more than one lesion, and some had none.

The research assistant recorded the radiologist's ratings for each processed digital image, as well as any other comments the radiologist made about the cases and/or digital processing algorithms. The research assistant then manually entered the data into an EXCEL spreadsheet (Microsoft, Redmond, Wash).

In sum, the 12 radiologists compared a total of eight processed images in each of the 28 cases (minus the seven images that were excluded) with the screen-film images. The total number of images viewed per radiologist was 441 (eight algorithms × 28 cases × two views = 448, minus the seven images that were not scored). The cases contained a total of 65 lesions, 29 that were histologically proved and 36 that were presumed benign.

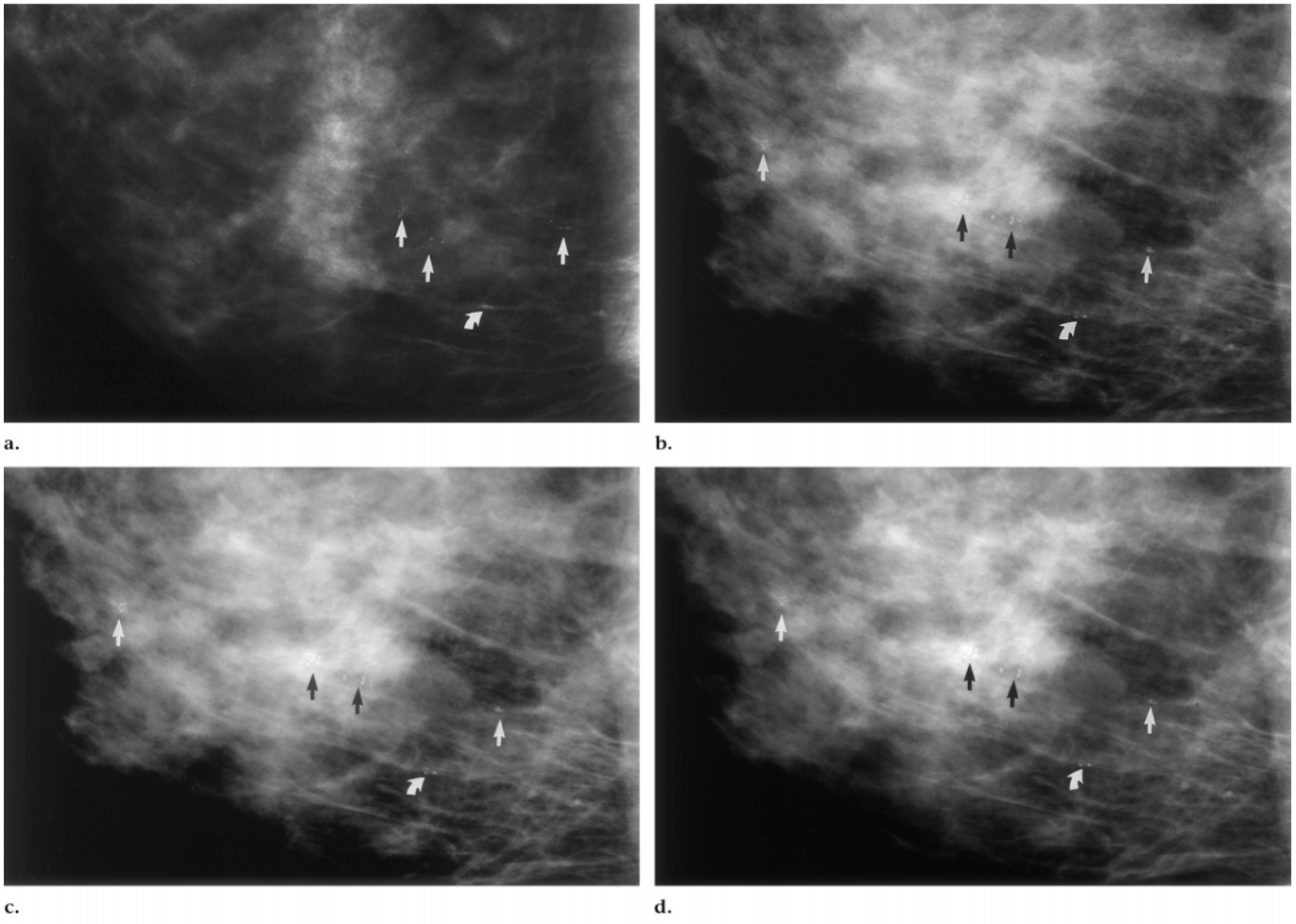


Figure 1. **(a)** Photographically magnified craniocaudal screen-film mammogram and **(b–d)** digital mammograms obtained with the Senographe 2000D unit (GE Medical Systems) in the same region of the same breast. Clustered calcifications (arrows) were localized by using needles and were proved to be atypical ductal hyperplasia at surgery. **(b)** Photographically magnified HIW-processed digital mammogram. **(c)** CLAHE-processed digital mammogram. **(d)** MIW-processed digital mammogram. MIW and automated windowing algorithms MMIW, HIW, and CLAHE somewhat compromise visibility of the skin for greater contrast in dense areas; all of these images scored better than the screen-film mammogram in the characterization of calcifications in this case. Algorithms designed to improve contrast while maintaining skin visibility were equivalent to (Trex processing) or worse than (Unsharp masking, MUSICA, and peripheral equalization) mammography. (Case provided by Daniel B. Kopans, MD, Massachusetts General Hospital, Boston.)

For the diagnostic task, there was one score per lesion on each of the two views for each of the eight algorithms minus the scores on the 18 lesions on the seven missing images for each of the 12 readers (65 lesions × two views × eight algorithms, minus 18 missing lesions that were not scored, × 12 readers for a total of 12,264 diagnostic scores collected).

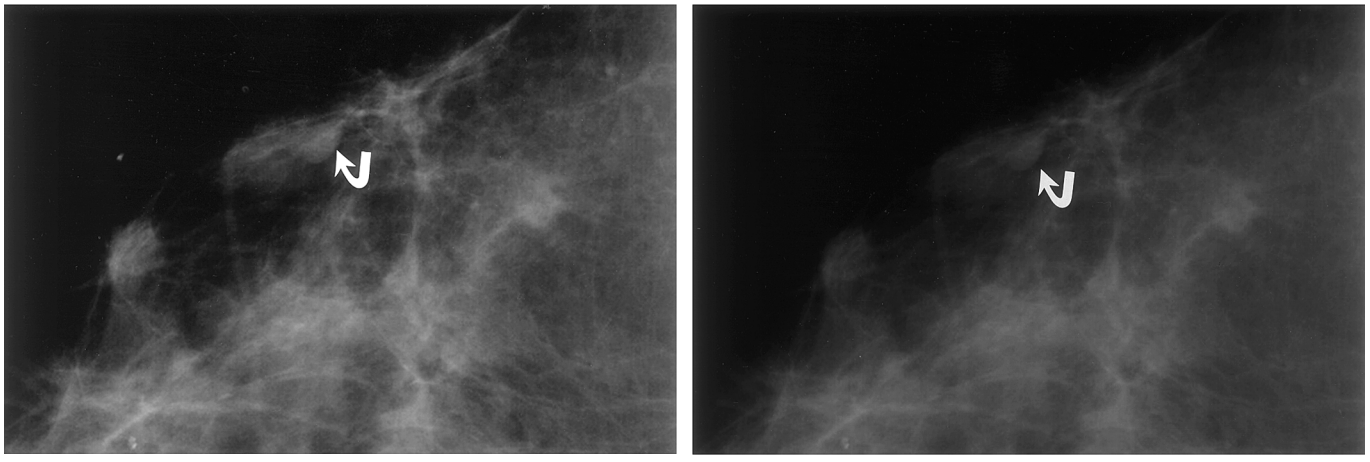
For the screening task, there were two views in the 28 cases with eight algorithms per case minus the seven missing images for 12 readers, or $[(28 \times 2 \times 8) - 7] \times 12$, for a total of 5,292 screening scores collected. Therefore, a total of 17,556 scores, or 12,264 + 5,292 scores, were requested from the 12 readers.

As some readers intentionally or accidentally failed to rate one or more lesions, the data set was incomplete. Some of the values were missed when a reader was unable to detect a lesion on either the screen-film mammogram or digital image and was, therefore, unable to rate it. Missing scores for lesions that were not visible on screen-film images were assigned a score of +2. To avoid possible bias toward digital images due to positioning differences, the two cases in which scores were resolved in this manner were excluded from the final analyses (although inclusion of the scores did not change results). Missing scores for lesions that were not visible on the digital image were assigned a score of −2. Cases affected by this assignment were retained in all analyses. Other missing scores were due to accidental oversights by the reader and the research assistant.

Statistical Methods

All primary and exploratory analyses were conducted separately for the three mammographic machines.

The primary analysis focused on the data for the diagnostic task and consisted of two parts. A mean for each processing method with the combination of lesion types was calculated by obtaining means for the reader, case, breast view, and lesion. Lesion types considered were calcifications and masses; masses with calcifications were classified as masses. Each of these 16 means (eight processing methods × two lesion types) was tested as being equal to zero,



a. b.

Figure 2. **(a)** Photographically magnified craniocaudal digital mammogram obtained with the SenoScan (Fischer Imaging) and processed with unsharp masking. Image shows a moderately well circumscribed mass (arrow) just below the skin in the far lateral portion of the breast. Mass was proved to be a simple cyst at ultrasonographically guided fine-needle aspiration. Because of its peripheral location, the lesion was not visible on images processed with MIW, MMIW, and HIW, which reduce visibility of subcutaneous detail to allow improved penetration and contrast in areas of highest density. **(b)** CLAHE-processed digital mammogram of the same area and mass (arrow).

which corresponded to a null hypothesis of no difference in radiologist's preference between the printed digital image and the screen-film mammogram. According to the Bonferroni technique for multiple comparisons, each test was evaluated at with $\alpha = .01/16 = .000625$, for an overall type I error rate of .01 for this set of tests.

In the second part of the primary analysis, model assumptions were verified, and the data were analyzed by using the analysis of variance (ANOVA) technique. The design for this two-way factorial repeated-measures ANOVA included lesion type, processing method, and their interaction. The test of method by lesion type interaction was conducted first, followed by step-down tests of the simple main effect of processing method within each lesion type. For each of the two lesion types, there were two (chosen from eight) times 28 pairwise comparisons among the digital processing methods for a total of 2×28, or 56 tests. According to the Bonferroni technique, each test was evaluated at the $\alpha = .04/56 = .000714285$ level, which resulted in an overall type I error rate of .04 for this set of tests. Note that the overall type I error rate for the complete primary analysis with each machine was .01 + .04, or .05.

The exploratory analysis of the screening task data mirrored the primary analysis. In the first part, a mean for each processing method by lesion type combination was calculated by averaging for reader, case, and breast view. Again, lesion types considered were calcifications and masses; masses with calcifications were classified as masses. Each of these 16 means (eight processing methods $\times$ two lesion types) was tested as being equal to zero, which corresponded to a null hypothesis of no difference in radiologist preference between the printed digital image and the screen-film mammogram with respect to breast cancer screening. Per the Bonferroni technique for multiple comparisons, each test was evaluated with $\alpha = .01/16 = .000625$, for an overall type I error rate of .01 for this set of tests. However, as this analysis was exploratory, *P* values must be interpreted as only descriptive statistics.

In the second part of the exploratory analysis, model assumptions were verified, and the data were analyzed by using the ANOVA technique. The design for this two-way factorial repeated-measures ANOVA included lesion type, processing method, and their interaction. The test of method by lesion type interaction was conducted first, followed by step-down tests of the simple main effect of processing method for each lesion type. For each of the two lesion types, there were two (chosen from eight) times 28 pairwise comparisons among the digital processing methods for a total of 2×28, or 56 tests. According to the Bonferroni multiple comparisons procedure, each test was evaluated at the $\alpha = .04/56 = .000714285$ level, which resulted in an overall type I error rate of .04 for this set of tests.

Finally, all means for method by lesion type were centered by subtracting the overall mean score for that machine. Centered means were computed for both the primary and exploratory analyses. To discourage comparison of the mean scores for the different mammographic machines, only the centered means are presented in the Results. However, note that all *P* values presented pertain to tests of the uncentered data.

All statistical analyses were performed by using SAS Software, version 6.12. (SAS Institute, Cary, NC.)

RESULTS

Primary Analysis: Diagnostic Mammographic Scores

Tables 4 and 5 show the radiologists' ratings of the digital processing algorithms with respect to the screen-film mammogram for the diagnostic mammographic tasks. Ratings are presented by machine type.

For each machine, there was statistically significant relationship between lesion type and preference of image processing algorithm for the lesion characterization and diagnostic mammographic tasks ($P = .0002$ for the Fischer Imaging unit, $P = .0024$ for the GE Medical Systems unit, and $P = .0338$ for the Trex unit). That is to say, for each machine, radiologists preferred different algorithms for the mass characterization and calcification characterization tasks.

Results with the Fischer Imaging unit.—For the diagnostic evaluation of masses

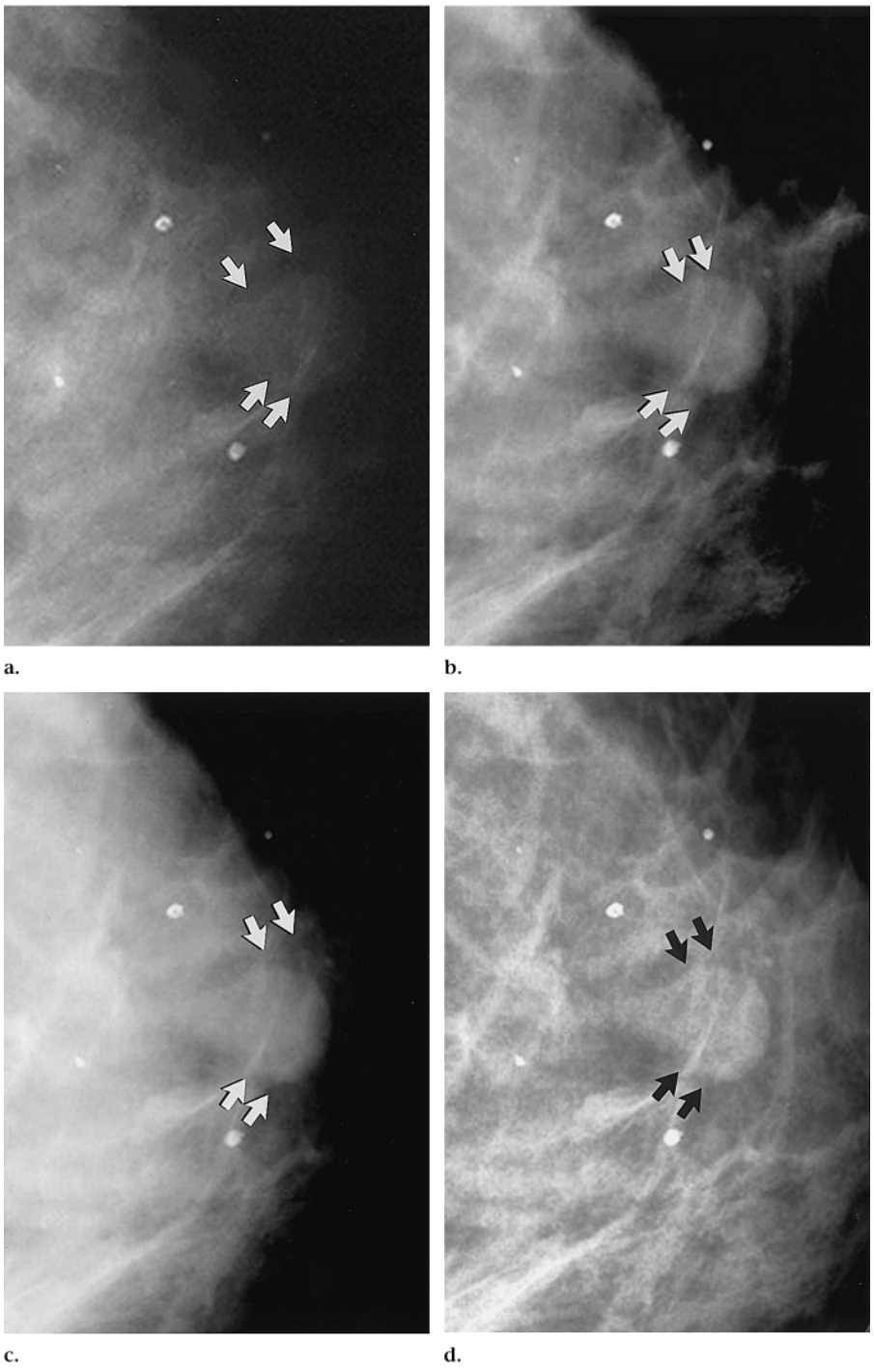


Figure 3. Images show a partially circumscribed, partially obscured nonpalpable mass (arrows) that had been visible for more than 1 year on mammograms and was, therefore, presumed to be benign. (Case provided by Emily Conant, MD, University of Pennsylvania, Philadelphia.) **(a)** Photographically magnified screen-film mammogram of the subareolar region. **(b)** Digital mammogram obtained with the Senographe 2000D unit (GE Medical Systems) and processed with unsharp masking, the algorithm study radiologists preferred for mass characterization on images obtained with this unit. Border conspicuity is improved compared with **a**. **(c)** MIW-processed digital mammogram. **(d)** MUSICA-processed digital mammogram.

(including masses with calcifications), all printed digital mammograms were preferred to the screen-film mammograms with all eight processing algorithms. Images processed with MUSICA, Trex processing, peripheral equalization, unsharp masking, and CLAHE were preferred significantly more at the $\alpha = .01/16 = .000625$ level. The machine-centered means for these algorithms were 0.37, 0.35, 0.32, 0.43 and 0.40, respectively.

For the diagnostic evaluation of calcifications, three of the eight printed processed digital mammograms—Trex, HIW, and MMIW images—were rated as being slightly better than or equivalent to the screen-film mammograms (machine-centered means of 0.15, 0.07, and 0.03, respectively). These differences were not statistically significant. The screen-film image was significantly favored over the digital images processed with MIW and peripheral equalization ($P < .000625$ or .01/16). The machine-centered means for these algorithms were −0.39, and −0.69, respectively.

Results with the GE Medical Systems unit.—For the mass diagnostic task, the digital image processed with unsharp masking was slightly but not significantly preferred to the screen-film image. The machine-centered mean score for unsharp masking was 0.18. The screen-film mammogram was significantly preferred over the images processed with the Trex algorithm at the $\alpha = .01/16 = .000625$ level; the machine-centered mean score for images processed with the Trex algorithm was −0.27.

For the calcifications diagnostic task, images processed with MIW, HIW, unsharp masking, and MMIW were all slightly preferred to the screen-film image. However, none of the preferences for digital processing algorithms were statistically significant. The machine-centered means for images processed with MIW, HIW, unsharp masking, and MMIW were 0.19, 0.34, 0.30 and 0.28, respectively. The screen-film mammogram was significantly preferred over the image processed with peripheral equalization at the $\alpha = .01/16 = .000625$ level; the machine-centered mean score for peripheral equalization was −0.48.

Results with the Trex unit.—For the mass diagnostic task, all processed digital images except those processed with MMIW were preferred to the screen-film mammogram; the images processed with HIW and Trex algorithms were significantly preferred at the $\alpha = .01/16 = .000625$ level. Machine-centered means for images processed with HIW and Trex algorithms were 0.53 and 0.57, respectively. The screen-film mammogram was preferred to the MMIW-processed image, but the difference was not statistically significant. The ma-

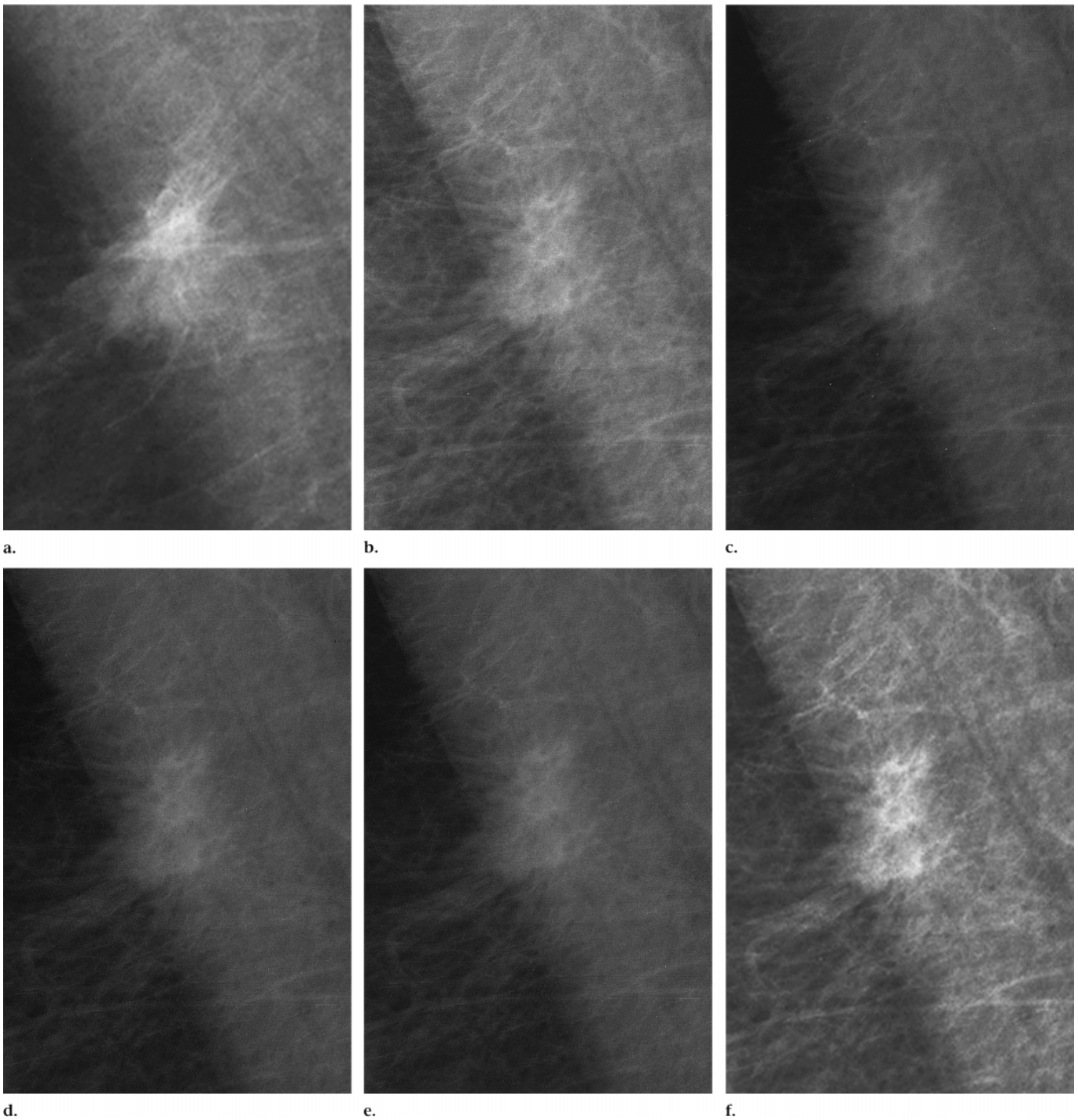


Figure 4. **(a)** Photographically magnified mediolateral oblique screen-film mammogram shows a spiculated mass in the axillary tail that was proved to be infiltrating lobular carcinoma and lobular carcinoma in situ at core biopsy and subsequent mastectomy. **(b–f)** Processed digital images are shown in the radiologists' order of preference in this case; all digital images had higher mean scores than that of the screen-film mammogram probably because spiculations on the anterior margin of the mass are more obvious on the digital images. **(b)** Digital mammogram obtained with the Digital Mammography System (Trex Medical) and processed with MUSICA. **(c)** Image processed with CLAHE. **(d)** Image processed with HIW. **(e)** Image processed with MIW. **(f)** Image processed with Trex algorithm. (Case provided by the Mark B. Williams, PhD, University of Virginia, Charlottesville, and Laurie L. Fajardo, MD, Johns Hopkins University, Baltimore, MD.)

chine-centered mean for MMIW was 0.17.

For the diagnostic evaluation of calcifications, the screen-film radiograph was significantly preferred over all eight processed digital images at the $\alpha = .01/16 = .000625$ level. The machine-centered mean scores ranged from 0.23 for the Trex processing algorithm to 0.75 for the peripheral equalization method. These results were statistically significant for all eight algorithms ($P < .01/16$ or .000625).

TABLE 4
Radiologist Preference for Mass Characterization

Algorithm	SenoScan (Fischer Imaging)	Senographe 2000D (GE Medical Systems)	Digital Mammography System (Trex Medical)
MUSICA	0.37 ± 0.34	0.03 ± 0.20	0.24 ± 0.14
Trex processing	0.35 ± 0.28	−0.27 ± 0.27	0.53 ± 0.19
MIW	0.21 ± 0.26	0.04 ± 0.22	0.42 ± 0.19
HIW	0.20 ± 0.34	−0.03 ± 0.26	0.57 ± 0.28
Peripheral equalization	0.32 ± 0.24	−0.02 ± 0.22	0.33 ± 0.30
Unsharp masking	0.43 ± 0.30*	0.18 ± 0.26	0.47 ± 0.23
CLAHE	0.40 ± 0.19*	−0.11 ± 0.18	0.44 ± 0.19
MMIW	0.13 ± 0.29	−0.09 ± 0.24	0.17 ± 0.16

Note.—For image processing algorithms applied to printed digital mammograms vs screen-film mammograms. Data are the mean score ± SD. Meaningful comparisons between machine types are not possible with these data.

* Mean score was significantly better ($P < .000625$ or 0.01/16).

TABLE 5
Radiologist Preference for Calcification Characterization

Algorithm	SenoScan (Fischer Imaging)	Senographe 2000D (GE Medical Systems)	Digital Mammography System (Trex Medical)
MUSICA	−0.34 ± 0.24	−0.02 ± 0.24	−0.32 ± 0.18*
Trex processing	0.15 ± 0.24	−0.23 ± 0.31	−0.23 ± 0.22
MIW	−0.39 ± 0.19*	0.19 ± 0.36	−0.44 ± 0.19*
HIW	0.03 ± 0.24	0.34 ± 0.40	−0.42 ± 0.29*
Peripheral equalization	−0.69 ± 0.38*	−0.48 ± 0.22*	−0.75 ± 0.16*
Unsharp masking	−0.11 ± 0.28	0.30 ± 0.44	−0.51 ± 0.11*
CLAHE	−0.31 ± 0.25	0.05 ± 0.32	−0.56 ± 0.17*
MMIW	0.07 ± 0.28	0.28 ± 0.38	−0.64 ± 0.19*

Note.—For image processing algorithms applied to printed digital mammograms vs screen-film mammograms. Data are the mean score ± SD. Meaningful comparisons between machine types are not possible with these data.

* Mean score was significantly worse ($P < .000625$ or 0.01/16).

TABLE 6
Radiologist Preference for Mass Screening Task

Algorithm	SenoScan (Fischer Imaging)	Senographe 2000D (GE Medical Systems)	Digital Mammography System (Trex Medical)
MUSICA	0.41 ± 0.47	0.44 ± 0.58	0.51 ± 0.49
Trex processing	0.84 ± 0.34	−0.48 ± 0.45*	0.91 ± 0.42
MIW	−0.14 ± 0.71	0.11 ± 0.39*	0.11 ± 0.37*
HIW	0.09 ± 0.57	0.26 ± 0.48	0.12 ± 0.45*
Peripheral equalization	0.07 ± 0.47	−0.24 ± 0.26*	−0.22 ± 0.43*
Unsharp masking	0.09 ± 0.78	−0.04 ± 0.49*	0.02 ± 0.38*
CLAHE	0.06 ± 0.47	−0.05 ± 0.39*	−0.11 ± 0.36*
MMIW	−0.50 ± 0.59*	0.01 ± 0.54*	−0.64 ± 0.37*

Note.—For image processing algorithms applied to printed digital mammograms vs screen-film mammograms. Data are the mean score ± SD. Meaningful comparisons between machine types are not possible with these data.

* Mean score was significantly worse ($P < .000625$ or 0.01/16).

Secondary Analysis: Overall Screening Score

Tables 6 and 7 show the radiologists' ratings of the digital processing algorithms with respect to the screen-film mammogram for the screening mammography tasks. Ratings are presented by machine type.

There was a relationship between lesion type and preferences of image processing algorithm with each machine for the lesion detection and screening mammographic score ($P = .0169$ for the Fischer Imaging unit, $P = .1025$ for the GE Medical Systems unit, and $P = .0165$ for the Trex unit). Since this was an exploratory analysis, P values may be interpreted only as descriptive statistics, and not as tests of significance.

Results with the Fischer Imaging unit.—For the detection of both masses and calcifications, only digital radiographs processed with the Trex algorithm were preferred to screen-film mammograms, although they were not strongly preferred. Machine-centered means for images processed Trex algorithms and obtained with the Fischer Imaging unit were 0.84 for masses and 1.0 for calcifications. The screen-film image was strongly preferred over the MMIW-processed images for both mass detection and calcification detection. Machine-centered means for MMIW were −0.5 and 1.0 for mass detection and calcification detection, respectively. The screen-film image was also strongly preferred over images processed with MIW, peripheral equalization, and unsharp masking for the detection of calcifications (machine-centered means of −0.27, −0.37, and −0.16, respectively). All tests were assessed at the $\alpha = .01/16 = .000625$ level.

Results with the GE Medical Systems unit.—For the detection of both masses and calcifications, the screen-film mammograms were preferred to the printed digital radiographs for all processing algorithms. For masses, the machine-centered mean scores ranged from 0.44 for the MUSICA algorithm to −0.48 for Trex processing. For calcifications, the machine-centered means ranged from 0.38 for MUSICA to −0.41 for peripheral equalization. All P values were less than .01/16 = .000625 except for MUSICA and HIW in the detection of masses and for MUSICA and MIW in the detection of calcifications.

Results with the Trex unit.—For the detection of masses, Trex processing was the only method preferred to the screen-film mammography, but it was not strongly preferred ($P > .000625$). The machine–centered mean for Trex processing in mass detection was 0.91. The screen-film mammogram was preferred to all other processed digital images for the detection of masses; centered means ranged from 0.91 for Trex processing to 0.64 for MMIW. The screen-film mammogram was preferred to all eight processed digital images for the detection of

calcifications; centered means ranged from 0.39 for MUSICA to 0.64 for CLAHE. *P* values were less than .01/16 = .000625 for all algorithms except the Trex and MUSICA algorithms for both lesion types.

TABLE 7
Radiologist Preference for Calcification Screening Task

Algorithm	SenoScan (Fischer Imaging)	Senographe 2000D (GE Medical Systems)	Digital Mammography System (Trex Medical)
MUSICA	0.32 ± 0.50	0.38 ± 0.47	0.39 ± 0.46
Trex processing	1.00 ± 0.34	0.13 ± 0.45*	0.28 ± 0.38
MIW	−0.27 ± 0.35*	0.31 ± 0.38	0.04 ± 0.40*
HIW	−0.10 ± 0.61	0.15 ± 0.48*	0.12 ± 0.34*
Peripheral equalization	−0.37 ± 0.47*	−0.41 ± 0.36*	−0.38 ± 0.40*
Unsharp masking	−0.16 ± 0.46*	−0.29 ± 0.51*	−0.38 ± 0.36*
CLAHE	−0.10 ± 0.57	0.11 ± 0.45*	−0.64 ± 0.29*
MMIW	−1.00 ± 0.34*	−0.29 ± 0.42*	−0.59 ± 0.45*

Note.—For image processing algorithms applied to printed digital mammograms vs screen-film mammograms. Data are the mean score ± SD. Meaningful comparisons between machine types are not possible with these data.

* Mean score was significantly worse ($P < .000625$ or 0.01/16).

DISCUSSION

Our results strongly indicate that radiologists prefer differently processed versions of the digital mammogram depending on the task, lesion type, and machine type. This finding suggests that digital mammograms would be best displayed by using monitor systems that allow flexibility and easy, quick access to differently processed versions of the images. If soft-copy interpretation is to become clinically practicable, ergonomic issues regarding image display with monitor systems must be overcome.

Undoubtedly, habit and experience influenced the preference of radiologists for screen-film images over processed digital images in many cases. A prior preference study (4), in which the investigators attempted to exactly match the appearance of the screen-film mammograms through MIW, showed that radiologists preferred digital mammography to screen-film imaging. Of course, such matching might not allow the full benefits of digital mammography to be realized.

Our study is limited by the fact that it was a preference study and not a quantitative measure of how well the radiologists performed. Radiologists provided their opinions on which images would improve their performance. Certainly they made educated guesses, but a performance study would have been better in the determination of how mammographic interpretation would be affected by image processing. This study is a good first step, however. A performance study would require many more cases and would have been too expensive and unwieldy if eight algorithms were tested. This experiment allows us to perform the next study as a performance study, with more cases and fewer algorithms to test.

In addition, this study is limited in that the diagnostic mammographic task did not include available compression and magnification views. However, since this limitation affected both modalities equally, it should not have substantially altered our results.

Clearly, the entire universe of image processing algorithms has not been tested. We chose those algorithms that were available to us, that were in clinical use, that we believed might have clinical utility, and about which we had expertise. Perhaps wavelets, derivatives, or an algorithm yet to be developed might have performed better than all of those tested and might have been better than screen-film mammography in all three tasks. In addition, a combination of algorithms (such as those available with a soft-copy display system) might allow for even better diagnostic performance and might have been most preferred by the radiologists.

In addition, since we included a small number of cases and since different lesions were imaged with each of the three systems, we believe that direct comparison of the results obtained with the three machines is not reasonable at this time. That is, we believe that the mean scores that the radiologists assigned to the various units for the various tasks should not be directly compared. We believe that we cannot justify statements about how the three units compare in the diagnosis or detection of masses or calcifications on the basis of these preliminary study findings alone. For example, clearly, the algorithms that were tested did not allow optimal characterization of calcifications with the digital images obtained with the Trex unit and did not allow optimal detection of calcifications and masses with the digital images obtained with the GE Medical Systems unit. We believe that these results are more reflective of the limitations of the algorithms tested than of the detectors themselves. In addition, some of the differences between the images could be attributed to the fact that the images obtained with the Trex unit were printed by using film, a printer, and a processor that were different from those used with the GE Medical Systems and Fischer Imaging units.

We omitted seven images from the total processed image data set because they could not be printed or because they were of poor quality. We believe that the size of any bias that was introduced into the study due to this factor was small, since the omitted images only represented 3.1% of the total sample prepared.

In fact, these results strongly suggest that the manufacturer of each digital mammographic system should determine which algorithms to use for optimal digital display with each mammographic task. These results will help in the guidance of those decisions. Clearly, some sort of objective performance measure (5,6) rather than an aesthetic assessment should be used by the manufacturers in guiding the selection of image processing algorithms. We believe that image processing might substantially enhance the achievable accuracy of digital mammography. Conversely, choices based on the production of digital mammograms that closely resemble screen-film radiographs might limit the results that can be achieved with this new technology.

Finally, we could not determine in this study whether other factors, such as breast density, patient age, location of the lesion within the breast, and other variables (such as soft-copy display) would influence radiologists' preferences regarding the algorithms. The role of these factors will have to be evaluated in future studies.

APPENDIX

MIW

For MIW, an experienced mammographic technologist manually set the intensity win-

dow for the digital mammograms displayed on a high-brightness (100-footlamberts) monitor (model 1654; Orwin Associates) with an Md5 Sun Display Card (Dome Imaging) and an UltraSparc computer (model 2200; Sun Microsystems). Both the monitor and display card has a display matrix of 2,048 × 2,560 pixels. The intensity windowing software was interactive, and the technologist can choose either a linear or asymmetric sigmoidal within-window intensity-mapping curve shape.

HIW

For HIW, the histogram for each mammogram in a study is automatically analyzed in terms of its peaks and troughs. All components of the breast tissue, such as the parenchyma, fatty areas, and skin edge portions, are recognized from these features of the histogram. With this method, contrast over the selected range of values of breast tissue is enhanced by means of simple intensity windowing.

MMIW

MMIW involves a combination of geometric (ie, intensity gradient–magnitude ridge traversal) and statistical (ie, Gaussian mixture modeling) techniques. This method is used to isolate the radiographically dense component on each mammogram, and, on the basis of the statistical characteristics of this isolated region, to set the parameters of an asymmetric sigmoidal intensity-mapping function.

CLAHE

CLAHE is a variant of adaptive histogram equalization. For adaptive histogram equalization, the histogram is calculated for the contextual region of a pixel, and a transformation provides the pixel with a new intensity that is proportional to its rank in the intensity histogram. It is designed to provide higher contrast for pixel intensities that occur more frequently and to provide a single displayed image in which contrast in all parts of the range of recorded intensities can be sensitively perceived. CLAHE limits the contrast increase factor produced with adaptive histogram equalization to a user-specified unit. The CLAHE parameter settings (clip, four; region size, 32) used in this study were based on previously published experimental results (5).

MUSICA

MUSICA processing is a multiscale wavelet-based contrast enhancement technique developed by Agfa. It involves variable enhancement of various spatial-scale components of the image and additive reconstruction. In our study, MUSICA processing was performed on an Agfa image-processing workstation. Three of its four image processing parameters, namely, edge contrast, latitude reduction, and noise reduction, were turned off by setting their levels to zero. The parameter for MUSICA was set to a maximum level of five.

Unsharp Masking

Unsharp masking is a technique used to crispen edges. A signal intensity proportional to the unsharp or low-passed filtered (blurred) version of the image is subtracted from the original image to yield a sharpened resultant image. The final image is produced by combining the original image (50% weighting) and the high-pass images (50% weighting). In our experiment, a region size of 600 × 600 pixels was used for the calculation of the low-pass image.

Peripheral Equalization

For peripheral equalization, thickness differences between the periphery of the breast and the center portions are smoothed so that the range of intensity values is accessible within the same narrow portion of the density look-up table. The thickness of the breast is approximated by using a smoothed version of the mammogram. with a resolution of about 3 mm. The perimeter of the breast is determined by using a simple threshold applied to the smoothed image and is expanded to a few millimeters outside the breast. Masking of pixels outside this area is applied to remove detector flat-fielding artifacts. The thickness effect is essentially removed by dividing the original image values by those on the smoothed image. The correction is applied only within 3 cm of the periphery of the breast, while areas within the center of the breast are left with their original values. A damping factor, which limits the magnitude of the correction, is applied to the pixels immediately adjacent to the edge of the breast to reduce ringing (6).

Trex Processing

The Trex processing method used in this study is the proprietary processing method applied as part of the Trex full-field Digital Mammography System. The algorithm is a weighted unsharp masking based on histogram data.

Author affiliations: The Departments of Radiology (E.D.P., B.M.H., R.E.J., M.E.B., S.R.A., M.P.B., R.M.), Biomedical Engineering (E.B.C., S.Z.), and Biostatistics (S.M., K.E.M., S.M.P.), University of North Carolina, Chapel Hill; the Breast Imaging Section, (E.C.) and the Department of Radiology, Physics Section (D.C.), University of Pennsylvania Medical Center, Philadelphia; the Department of Radiology, Johns Hopkins University, Baltimore, Md (L.L.F.); the Division of Breast Imaging, Mount Sinai Medical Center, New York, NY (S.A.F.); the Department of Radiology, University of Chicago, Ill (R.M.N.); the Departments of Medical Biophysics and Medical Imaging (M.J.Y., D.B.P.) and Radiology (R.S.), University of Toronto, Sunnybrook Health Sciences, Ontario, Canada; the Department of Radiology, University of Virginia Health Sciences Center, Charlottesville, (M.B.W.); Hillsborough, NC (L.T.N.); the Department of Radiology, Thomas Jefferson University Hospital, Philadelphia, Pa (A.D.A.M.); Kalamazoo Radiology, Mich (A.V.); the Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Mass (P.J.K.); the Department of Radiology, Duke University, Durham, NC (E.R., M.S.S., R.W., M.W.); the Department of Radiology, Massachusetts General Hospital, Boston (D.B.K., R.H.M.); the Department of Medical Imaging, Mount Sinai Hospital, Toronto, Ontario, Canada (R.J.); and the Department of Radiology, Good Samaritan Hospital Medical Center, West Islap, NY (M.S.).

Other Co-authors and Members of the International Digital Mammography Development Group: M. Patricia Braeuning, MD, Robert McLelland, MD, Stephen M. Pizer, PhD, Marylee E. Brown, BA, University of North Carolina, Chapel Hill; Eric Rosen, MD, Mary Scott Soo, MD, Margaret Williford, MD, Duke University, Durham, NC; Loren T. Niklason, PhD, Hillsborough, NC; Andrew D. A. Maidment, PhD, Thomas Jefferson University Hospital, Philadelphia, Pa; Andrei Vermont, MD, Kalamazoo Radiology, Mich; Phyllis J. Kornguth, MD, PhD, Brigham and Women's Hospital, Boston, Mass; Daniel B. Kopans, MD, Richard H. Moore, BA, Massachusetts General Hospital, Boston; Dev Chakraborty, PhD, University of Pennsylvania, Philadelphia; Roberta Jong, MD, Mount Sinai Hospital, Toronto, Ontario, Canada; Rene Shumak, MD, University of Toronto, Ontario, Canada; Melinda Staiger, MD, Good Samaritan Hospital Medical Center, West Islap, NY; and Donald B. Plewes, PhD, Sunnybrook Health Sciences, Toronto, Ontario, Canada.

Funding/Support: E.D.P. supported in part by National Cancer Institute, National Institutes of Health grant number RO1 CA60193-05; Office of Women's Health, Department of Health and Human Services grant number 282-97-0078; and U.S. Army Medical Research and Material Command, grant #DAMD 17-94-J-4345. M.J.Y. supported by Canadian Breast Cancer Research Initiative, grant number 7289. S.A.F. supported by National Cancer Institute, National Institutes of Health grant number 1RO1CA60192. D.C. supported by National Cancer Institute, National Institutes of Health grant number RO1-CA75145-01A1 and Office of Women's Health, Department of Health and Human Services grant number 282-97-0077. R.M.N. supported by National Cancer Institute, National Institutes of Health grant number RO1-CA60187. The University of North Carolina Department of Radiology is being paid by Fischer Imaging for E.D.P. to assist in the design and implementation of the Food and Drug Administration approval trial for the SenoScan digital mammography system. D.P. owns stock in General Electric.

Author contributions: Guarantor of integrity of entire study, E.D.P.; study concepts, E.D.P., B.M.H., R.E.J., M.J.Y., R.M.N.; study design, E.D.P., B.M.H., K.M., R.E.J., M.J.Y., R.M.N.; definition of intellectual content, all co-authors; literature research, E.D.P.; experimental studies, E.D.P., E.B.C., B.M.H., M.W., S.R.A., S.Z., L.T.N., A.D.A.M., M.P.B., R.M., A.V., P.J.K., E.R., M.S.S., R.W., M.W., E.C., L.L.F., D.B.K., D.C.; data acquisition, M.P.B., R.M., A.V., P.J.K., E.R., M.S.S., R.W., M.W., E.B.C., S.Z.; data analysis, S.M.,

K.E.M., E.D.P.; statistical analysis, S.M., K.E.M.; manuscript preparation, E.D.P.; manuscript editing and review, all co-authors.

Acknowledgments: The authors gratefully acknowledge the contributions of the following individuals in support of this work: Faina Shtern of the Department of Radiology, Beth Israel Deaconess Medical Center, Boston, Mass; Michelle Picarro, Michael Tesic, Ruth Grafton, Pat Campbell, and Morgan Nields of Fischer Imaging, Denver, Colo; Beale Opsahl-Ong of GE Corporate Research and Development, Milwaukee, Wis; Lim Cheung and Richard Bird of Trex Medical, Long Island, NY; Art Haus, Theresa Bogucki, and Jeff Byng of Eastman Kodak, Rochester, NY; Willem Van Reit, Charles Augello, and John Landry of Agfa Division of Bayer, Ridgefield, NJ; Jenny Harrison, Patty Barbour, Lisa Quamme, Kerrie Kurgat, Mark Kramer, Sheldon Earp, Anna Cleveland, Sanjay Sthapit, and Joseph K.T. Lee of the University of North Carolina, Chapel Hill; Wendy Kurutz of Good Samaritan Hospital, West Islap, NY; Denise McDonald of the University of Pennsylvania, Philadelphia; Anoma Gunasekara of the University of Toronto, Ontario, Canada; Lisa Sparacino of Thomas Jefferson University, Philadelphia, Pa; Jayne Cormier of Massachusetts General Hospital, Boston; Claire Poyet and Karen Seaton of Wake Radiology, Raleigh, NC; Elizabeth Eagle and Rebecca Kennedy of Greensboro Radiology, NC; and Mary Baldwin of the University of Virginia, Charlottesville. This work would not have been possible without the generous support of Fischer Imaging, GE Medical Systems, Trex Medical, Eastman Kodak, and the Agfa Division of Bayer. The companies provided equipment and support services for the performance of this study.

References

1. Weibrecht M, Spekowius G, Quadflieg P, Blume H. Image quality assessment of monochrome monitors for medical soft copy display. SPIE 1997; 3031:232–244.
2. Shtern F, Winfield D, eds. Report of the working group on digital mammography: digital displays and workstation design. Washington, DC: Public Health Service Office of Women's Health and National Cancer Institute, 1998.
3. National Electrical Manufacturers Association Web site. Available at: *http://www.nema.org*. Accessed October 1998.
4. Pisano ED, Chandramouli J, Hemminger BM, et al. Does intensity windowing improve the detection of simulated calcifications in dense mammograms? J Digit Imaging 1997; 10:79–84.
5. Pizer SM. Psychovisual issues in the display of medical images. In: Hoehne KH, ed. Pictorial information systems in medicine. Berlin, Germany: Springer-Verlag, 1985; 211–234.
6. Byng JW, Critten JP, Yaffe MJ. Thickness equalization processing for mammographic images. Radiology 1997; 203:564–568.