

Quality control for digital mammography in the ACRIN DMIST trial: Part I

Aili K. Bloomquist, Martin J. Yaffe, Etta D. Pisano, R. Edward Hendrick, Gordon E. Mawdsley, Stewart Bright, Sam Z. Shen, Mahadevappa Mahesh, Edward L. Nickoloff, Richard C. Fleischman, Mark B. Williams, Andrew D. A. Maidment, Daniel J. Beideck, Joseph Och, and J. A. Seibert

Citation: *Medical Physics* **33**, 719 (2006); doi: 10.1118/1.2163407

View online: <http://dx.doi.org/10.1118/1.2163407>

View Table of Contents: <http://scitation.aip.org/content/aapm/journal/medphys/33/3?ver=pdfcov>

Published by the American Association of Physicists in Medicine

Articles you may be interested in

[Method of measuring NEQ as a quality control metric for digital mammography](#)
Med. Phys. **41**, 031905 (2014); 10.1118/1.4865175

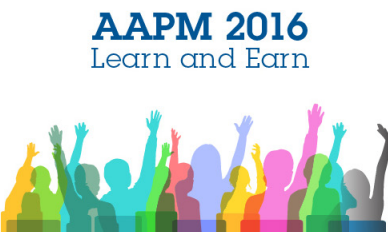

[Contrast sensitivity of digital imaging display systems: Contrast threshold dependency on object type and implications for monitor quality assurance and quality control in PACS](#)
Med. Phys. **36**, 3682 (2009); 10.1118/1.3173816

[Physical characteristics of five clinical systems for digital mammography](#)
Med. Phys. **34**, 2730 (2007); 10.1118/1.2742498

[Quality control for digital mammography: Part II recommendations from the ACRIN DMIST trial](#)
Med. Phys. **33**, 737 (2006); 10.1118/1.2164067

[Validation of MTF measurement for digital mammography quality control](#)
Med. Phys. **32**, 1684 (2005); 10.1118/1.1921667

Related AAPM Publications

<p>Educational Lectures</p> <p>Don't miss these fascinating in-booth speakers. Lectures will be held throughout the show during exhibit hours only, in booth #4001.</p> <p>Joe Ting, PhD Utilizing EPID for stereotactic cone commissioning and verification in RIT</p> <p>Sam Hancock, PhD Isocenter optimization tools for LINAC-based SRS/SBRT</p>	 <p>AAPM 2016 Learn and Earn</p>	<p>Users Meeting</p> <p>Enjoy some delicious dessert while you learn and earn 2 CAMPEP credit hours at our Users Meeting.</p> <p>Location . . . Marriott Marquis, Washington, DC</p> <p>Date Sunday, July 31</p> <p>Time 7-9 PM</p>	<p>Visit us at AAPM Booth #4001</p>  <p>call or visit 719.590.1077 • radimage.com</p> <p><small>© 2016 RadImage Imaging Technology, Inc. 2016-07-26</small></p>
--	--	--	---

Quality control for digital mammography in the ACRIN DMIST trial: Part I

Aili K. Bloomquist and Martin J. Yaffe

*Imaging Research Program, Sunnybrook and Women's College Health Sciences Centre,
2075 Bayview Avenue, Toronto, Ontario M4N 3M5, Canada*

Etta D. Pisano

*Department of Radiology and Lineberger Comprehensive Cancer Center, University of North Carolina at
Chapel Hill, Chapel Hill, North Carolina 27599-7295*

R. Edward Hendrick

*Lynn Sage Comprehensive Breast Center and Department of Radiology, Northwestern University's Feinberg
Medical School, Galter Pavilion, 13th Floor, 251 E. Huron Street, Chicago, Illinois 60610*

Gordon E. Mawdsley, Stewart Bright, and Sam Z. Shen

*Imaging Research Program, Sunnybrook and Women's College Health Sciences Centre,
2075 Bayview Avenue, Toronto, Ontario M4N 3M5, Canada*

Mahadevappa Mahesh

*The Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University,
JHOC Suite 4235, 601 N. Caroline Street, Baltimore, Maryland 21287-0856*

Edward L. Nickoloff

*Department of Radiology, Columbia University, College of Physicians & Surgeons,
177 Fort Washington Avenue, New York, New York 10032*

Richard C. Fleischman

*Department of Medical Physics, Memorial Sloan Kettering Cancer Center, 1275 York Avenue,
New York, New York 10021*

Mark B. Williams

*Radiology, Biomedical Engineering and Physics, University of Virginia, Box 801339,
Charlottesville, Virginia, 22908*

Andrew D. A. Maidment

*Department of Radiology, Hospital of the University of Pennsylvania, 1 Silverstein, 3400 Spruce Street,
Philadelphia, Pennsylvania 19104*

Daniel J. Beideck

*Radiology Department, Fletcher Allen Health Care, University of Vermont, 111 Colchester Avenue,
Burlington, Vermont 05401*

Joseph Och

Allegheny General Hospital, Medical Imaging, 320 East North Avenue, Pittsburgh, Pennsylvania 15212

J. A. Seibert

*Department of Radiology and Biomedical Engineering, University of California Davis,
Sacramento, California 95817*

(Received 14 March 2005; revised 1 December 2005; accepted for publication 5 December 2005;
published 23 February 2006)

The Digital Mammography Imaging Screening Trial, conducted by the American College of Radiology Imaging Network, is a clinical trial designed to compare the accuracy of full-field digital mammography (FFDM) versus screen-film mammography in a screening population. Five FFDM systems from four manufacturers (Fischer, Fuji, General Electric, and Lorad) were employed in the study at 35 clinical sites. A core physics team devised and implemented tests to evaluate these systems. A detailed description of physics and quality control tests is presented, including estimates of: mean glandular dose, modulation transfer function (MTF), 2D noise power spectra, and signal-to-noise ratio (SNR). The mean glandular doses for the standard breast ranged from 0.79 to 2.98 mGy, with 1.62 mGy being the average across all units and machine types. For the five systems evaluated, the MTF dropped to 50% at markedly different percentages (22% to 87%) of the Nyquist limit, indicating that factors other than detector element (del) size have an important effect on spatial resolution. Noise power spectra and SNR were measured; however, we found that it was difficult to standardize and compare these between units. For each machine type, the performance as measured by the tests was very consistent, and no predictive benefit was seen for many of the tests during the 2-year period of the trial. It was found that, after verification of proper operation during acceptance testing, if systems failed they generally did so suddenly rather than through gradual deterioration of performance. Because of the relatively short duration of this study further,

investigation of the long-term failure characteristics of these systems is advisable. © 2006 American Association of Physicists in Medicine. [DOI: 10.1118/1.2163407]

Key words: digital mammography, quality control, image quality

I. INTRODUCTION

The Digital Mammography Imaging Screening Trial (DMIST) is being conducted to compare the diagnostic accuracy of digital versus screen-film mammography in a screening population of 49529 women enrolled to receive both digital and screen-film mammograms at 35 clinical sites.¹ The trial is funded by the National Cancer Institute (NCI) and is being managed by the American College of Radiology Imaging Network (ACRIN).

This report presents a review of the quality control (QC) procedures used in DMIST and a summary of the test results.

It must be emphasized that most of the units were not mature products. During DMIST, one system became obsolete, and software on most systems underwent revisions. A number of detectors were replaced due to design problems, which have since been overcome, and maintenance procedures for the equipment have been modified so that a number of problems found during the trial will no longer appear.

An extensive set of physics acceptance and QC tests was developed to evaluate all digital mammography units involved in the trial. The protocol and initial pass/fail criteria were established at the outset of the trial and designed so that, as much as possible, tests could be applied generically among the different digital mammography systems. Whenever appropriate, the design of the tests was based on the American College of Radiology's Quality Control Manual,² and pass-fail criteria taken from the Mammography Quality Standards Act (MQSA) requirements for screen-film systems.³ It was also necessary to incorporate tests as prescribed by individual manufacturers of digital mammography systems used in DMIST. In addition, test input from the International Digital Mammography Development Group (IDMDG)⁴ was used and assumptions with regard to failure modes of the various systems guided test development. These tests were developed before the American Association of Physicists in Medicine's Task Group 18 (AAPM TG-18) developed tests for monitors.

Some tests and criteria were refined through the trial as QC problems became evident. Limits were set to be consistent with equipment design and to ensure that there would not be a high number of failures due to factors over which the users had no control.

Three groups of tests evaluated image quality from acquisition to display. The first group of tests characterized x-ray equipment operation and detector performance. A second group measured a number of image quality and dose parameters including: imaging and scoring the American College of Radiology mammography accreditation phantom (MAP); measuring entrance exposure and mean glandular dose (MGD); evaluation of a phantom developed especially for digital mammography; evaluation of image artifacts, noise

power spectra, modulation transfer functions (MTFs), linearity, and reproducibility. The third group focused on the image display systems; these tests evaluated monitor and printer calibration.

The initial physics testing for the study at all 35 DMIST sites was performed by a single central physicist (A.K.B.) from the DMIST QC core group, who carried out a series of tests on the digital equipment and trained the local (on-site) physicist and the technologists in the QC procedures. The screening sites were then tested at 6-month intervals (the "testing period") by the local physicist.

In the study design, the established protocol was maintained as closely as possible to allow monitoring of performance over time and across types of equipment. A key goal was to establish those tests that could detect image quality problems or predict failure in order to assist in the development of a rational QC program for digital mammography.

For some equipment, multiple testing was performed, as required following hardware or software upgrades or repairs. For other systems (e.g., those that entered the trial near its conclusion), only one test point was available. The performance of the screen-film mammography (SFM) systems used in the trial was monitored through the QC documentation required under MQSA.

Here, in Part I of this report, the unique QC tests are described and basic results from DMIST are presented. See Ref. 28 for more detailed results. Objectives and pass/fail criteria for the tests are given in Part II of this report.⁵ Examples of the types of problems uncovered by these tests during evaluation of the mammography units are given. Based on this experience, recommendations for an effective and practical QC program applicable to digital mammography will be presented in Part II.

II. DIGITAL MAMMOGRAPHY EQUIPMENT

Five different full-field digital mammography (FFDM) systems from four manufacturers were used in the DMIST trial. These included the Senoscan (Fischer Imaging Corporation, Denver, CO), referred to as "Fischer"; FCR 5000 MA (Fujifilm Medical Systems, Japan) "Fuji"; Senographe 2000D (GE Medical Systems, Waukesha, WI), "GE"; Lorad Digital Breast Imager FFDM System (Hologic, Inc., Danbury, CT), "Lorad DBI" and Lorad Selenia FFDM System (Hologic, Inc., Danbury, CT), "Lorad Selenia." These systems have been described elsewhere⁶⁻¹¹ and a summary comparison of the technologies tested is given in Table I.

Four different laser printers from three manufacturers were used to print the digital mammograms acquired on the Fuji CR system, the Lorad DBI system, and the Lorad Selenia system for interpretation by the radiologist. These included the Agfa Drystar 4500 (Agfa Medical Imaging, Wilm-

TABLE I. Summary of the digital mammography system technologies tested in the study (del=detector element, 1 k=1024).

System	Units #	Del size μm	Image matrix size K \times K	Bit depth	Technology	Grid
Fischer	7	54	4.0 \times 5.5	12	CsI, CCDs, slot-scan	No
Fuji	6	50	4.6 \times 5.8	10 (log)	dual-side CR	Yes
GE	20	100	1.9 \times 2.3	14	CsI on α -Si	Yes
Lorad DBI	6	40	4.7 \times 6.3	14	CsI, CCDs	Yes
Lorad Selenia	6	70	3.3 \times 4.0	14	α -Se	Yes

ington, DE), Fuji FM-DPL (Fujifilm Medical Systems), and Kodak DryView 8610 (Eastman Kodak Company, Health Imaging Division, Rochester, NY), which are dry-process printers, and the Agfa Scopix LR5200 (Agfa Medical Imaging, Wilmington, DE), which is a wet-process printer. A summary comparison of the printers is given in Table II.

III. TEST OBJECTS USED

A. Mammographic accreditation phantom (MAP)

The MAP used in the American College of Radiology (ACR) Mammography Accreditation Program² and mandated by MQSA was used as a reference for subjective evaluation of image quality. All sites already had and were familiar with the phantom from its use in the QC program for SFM systems.

B. "Misty" phantom

Misty is a phantom developed for the evaluation of digital mammography systems by the IDMDG. It has been described in detail elsewhere.¹²⁻¹⁴ It is 5.5 cm thick and composed of polymethyl methacrylate (PMMA) with a mercury-intensified radiographic overlay.¹⁵ It contains several test objects including low-contrast resolution line-pair patterns, rulers for measuring tissue excluded from the image at the chest wall, a central region containing no structural detail that can be used to calculate the noise power spectrum, and a step wedge for evaluating contrast (Fig. 1). Conspicuity is also measured using groups of stars of decreasing size ranging from 700 to 170 microns [Fig. 2(a), distance "d" in Fig. 2(b)]; the stars are evaluated over two thicknesses of PMMA, 2.8 and 5.4 cm.

Images of this phantom were acquired on all systems tested using clinically relevant exposure techniques for a 5.5 cm thick breast.

C. Uniform phantom

A 4.0 cm thick uniform slab of PMMA covering the entire detector was used for artifact evaluation, noise analysis, and routine weekly QC.

D. MTF tool

A special test tool was used to measure the system modulation transfer function (MTF). The tool consists of a square of either niobium (with ground edges) or brass (precision punched) mounted on a larger square of 1 mm thick aluminum (see Fig. 3).

E. Distortion tool

A printed circuit board with a grid of thin copper lines, forming 2 \times 2 cm squares angled at 45 deg (Fig. 4), was used to evaluate the imaging chain for distortion.

F. SMPTE pattern

A modified version of the Society of Motion Picture and Television Engineers (SMPTE) medical diagnostic imaging test pattern¹⁶ was used to assess the calibration and spatial resolution of monitors used to read the digital mammograms (Fig. 5). This pattern was modified so that for each type of digital acquisition system, the image size (number of pixels in length and width) of the pattern matched the image size of

TABLE II. Summary of the printer technologies used in the study.

System	Units #	Pixel size (μm)	Film size (cm)	Matrix size (k)	Bit depth	Technology
Agfa 4500	1	50	20.3 \times 25.7	3.8 \times 4.9	12	Solid state, heat-sensitive film
Agfa 5200	6	40	20.3 \times 25.7	4.9 \times 5.8	8	He-Ne Laser, wet processing
Fuji FM-DPL	6	50	25.7 \times 36.4	5.1 \times 7.3	12	Laser, dry processing
Kodak 8610	7	40	20.3 \times 25.7	5.0 \times 6.2	12	Laser, dry processing

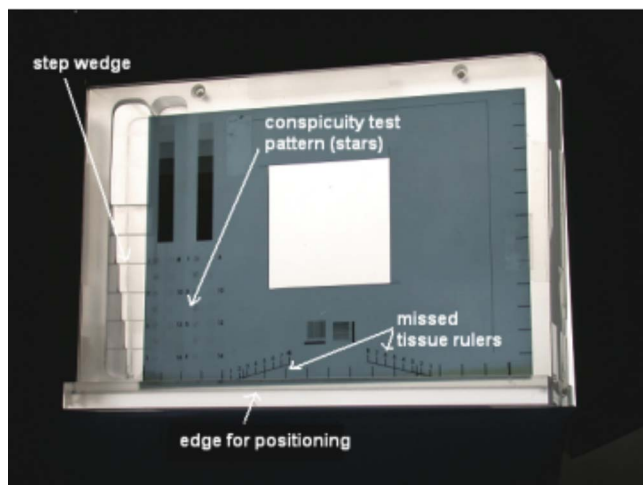


FIG. 1. Photograph of Misty phantom, with different components used in DMIST indicated.

the acquisition device, and the finest level of resolution was created by setting one pixel “on” (maximum level) and the adjacent pixel “off” (0).

IV. TESTS PERFORMED

The testing sequence was designed to attempt to maximize efficiency, and, where possible, combined multiple tests in a single exposure. Tests which did not differ markedly from those required for SFM by MQSA are not described in detail. These include: unit evaluation, collimation, kV accuracy, and half-value layer measurement. We discuss the results in each section, and present recommendations in the accompanying article (Part II).

A. X-ray production and physical safety

All dosimeters and kV meters used in this study were calibrated to meet MQSA requirements.

1. Unit evaluation

a. Method. In testing similar to that for SFM, both the physicist and technologist performed an overall inspection of the system for mechanical safety, operation of controls, and compression adequacy.

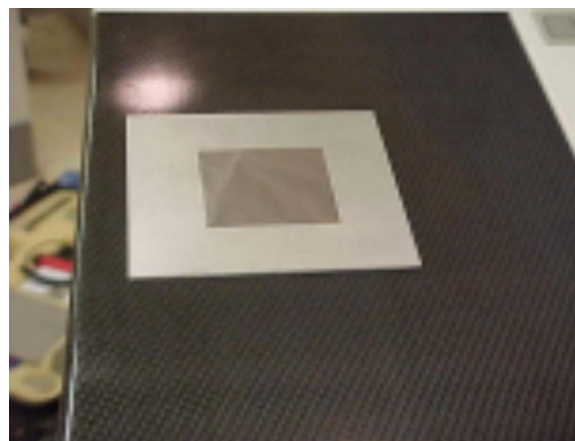


FIG. 3. MTF tool.

b. Results and discussion. From 11697 technologist's daily unit evaluations recorded in the DMIST QC database, there were 62 problems reported. The most frequently recorded problems were with the viewing conditions. There were 29 instances where viewing conditions were recorded as unacceptable and 20 instances where monitors were recorded as not being clean. Other problems include cracks in the compression paddle (8), crimped hoses or cables (2), unit cleanliness (2), and overall unit integrity (1).

In 146 physicist's unit evaluations, there were 12 problems noted. The only documented problems were missing technique charts (5), inaccurate thickness scale readouts (4), and maximum compression in the initial power drive mode pressure exceeding 20 daN (3). The maximum recorded compression force was 21 daN, which is probably within the measurement error of the various compression tools used.

The relatively low incidence of deficiencies recorded may be due to the fact that QC records were monitored monthly and that the technologists involved in the study were highly motivated.

2. Imaging plate fogging (CR Only)

a. Method. To evaluate if the imaging plates would be fogged by radiation in their storage location in the mammography room, a coin was taped to the “tube” side of an imaging plate cassette and left in the location used to store unex-

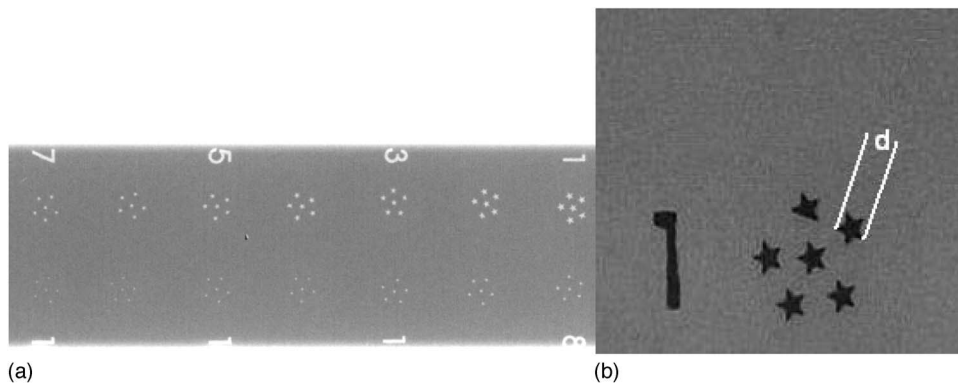


FIG. 2. (a) Region of interest from radiograph of a Misty phantom showing the 14 groups of stars. One star in each group is missing a point. (b) Photograph of single group of stars from Misty phantom. The top star in this group is missing a point. The star size “d” ranges from 175 to 700 microns.

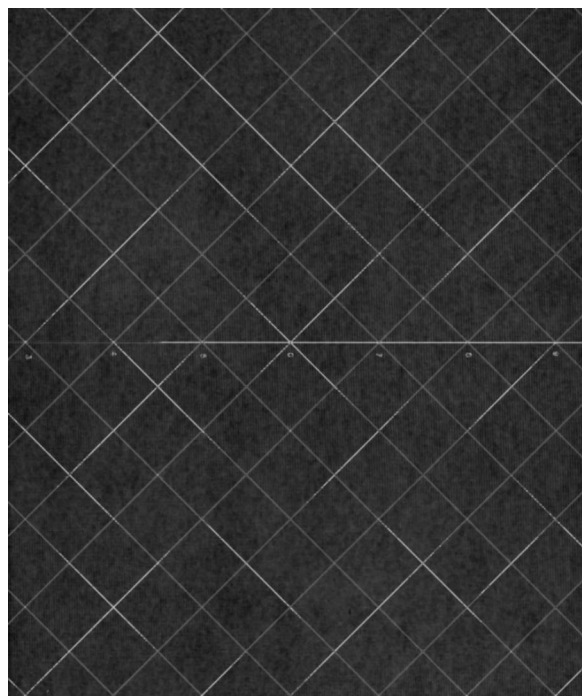


FIG. 4. Radiograph of the distortion test tool. The lines are 2 cm apart, copper on printed circuit board.

posed cassettes for the duration of the physics testing. If the image of the coin was not visible, fogging was determined to be negligible.

b. Results and discussion. No evidence of imaging plate fogging was seen on any of the 10 units that underwent testing.

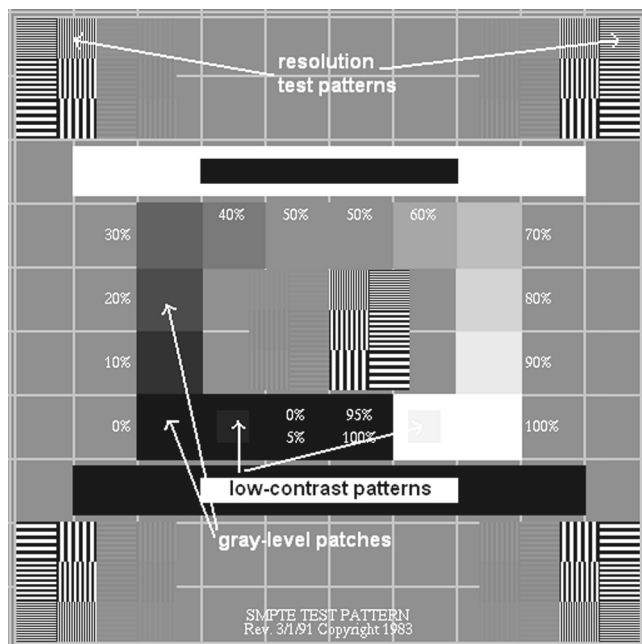


FIG. 5. SMPTE pattern.

3. Collimation and alignment

a. Methods. Radiographic rulers accurate to 1 mm were used to define the field indicated by the machine (positioning light or other indicator). An identifying device (radiographic ruler or paper clip) was placed at the inner lip of the compression plate located 4.5 cm above the breast support table. An image taken with a mammographic screen-film cassette was used to identify the actual edges of the x-ray field. Collimation requirements were similar to those specified in the ACR QC program for SFM.² A measurement tool in the Misty test phantom allowed assessment of tissue missed at the chest wall. This measurement was only carried out on the most frequently used target-filter combination and image-receptor size.

b. Results and discussion. Congruence between the x-ray field and the field indicator generally met MQSA requirements, especially for the various mammography units used with the Fuji system, and the GE and Lorad DBI systems. Problems were sometimes encountered with the Fischer system, where the field is indicated by lines printed on the breast support surface. Congruence problems in the left-right direction on Fischer digital systems (failure rate of 20% in 61 measurements) could sometimes be corrected by adjusting the synchronization between the scanning motion and the x-ray exposure.

Congruence between the x-ray field and the image receptor was also generally good. Once again, problems were encountered on some Fischer systems (failure rate of 18%). Achieving the MQSA suggested limit of 2% of the source-to-image distance was problematic on the Fischer system. If image acquisition began too early, the leading edge of the image would be underexposed, as the tube output had not yet stabilized. Similarly, if image acquisition commenced too late, there would be an increased duration of x-ray exposure without image data being acquired. For the Lorad Selenia system, in some cases the evaluation was done on hard copy, and the failure rate of 8% in 24 measurements and the large maximum error of 26 mm were due to the film printing algorithm (since corrected), which centered and cropped the image, removing part of the chest-wall side of the breast from the image.

The congruence between the chest-wall edge of the compression paddle and the edge of the image receptor was excellent for the various units used with the Fuji system, and for the GE system, with no failures in 74 and 120 measurements, respectively. The failures on the Lorad DBI (10% in 10 measurements) and Selenia system (4% in 24 measurements) were due to the mispositioning of the compression paddle and were easily corrected. Missing tissue at the chest wall exceeded the 7 mm limit declared at the beginning of the study for some of the Fischer, Lorad DBI, and Lorad Selenia units (failure rates of 31% in 35 measurements, 75% in 8 measurements, and 15% in 13 measurements, respectively) as well as some of the screen-film units being used with the Fuji system (11% in 28 measurements).

For the Fischer system, the larger amount of missed tissue is design related and is due to the need to allow enough

space between the scanning arm and the tabletop cover to prevent the cover from interfering with the scanning arm motion. The breast support was frequently removed by the service engineer, and could be repositioned incorrectly, causing more tissue to be missed. For the Lorad DBI system, the amount of missed tissue is due to a combination of the bucky design and the positioning of the compression paddle too far forward, such that it pushed the lip of the test phantom away from the chest-wall edge of the detector. Likewise, the position of the compression paddle may have contributed to the Lorad Selenia failures. In our experience, missed-tissue measurements on screen-film units in the field are occasionally found to fall outside the 7 mm limit.

4. kV Accuracy and reproducibility

a. Method. The kilovoltage output accuracy and reproducibility were evaluated in the same manner as for screen-film systems, across the range of clinically relevant kV settings. The scanning geometry and unusual beam spectra (Tungsten anode, aluminum filter) of the Fischer system made reliable kV estimation with a noninvasive meter difficult. For these units, kV measurements were obtained by the service person using a volt meter connected to a voltage divider contained within the generator.

b. Results and discussion. For each measured kV, the absolute percent difference between the nominal and measured kV was calculated. In data from 137 testing inspections, there was only one test instance where the measured output kV was beyond 5% of the nominal kV. In addition, there were no cases where the coefficient of variation (COV) among repeated kV measurements was greater than 0.05.

5. Tube output, linearity, output rate, and reproducibility

a. Method. To characterize the generator performance and enable the estimation of breast entrance exposure, the exposure output in mR/mAs was measured over a range of clinically relevant kV settings and for the available x-ray target/beam filter combinations. In addition, the output rate was measured at 28 kV for the molybdenum/molybdenum anode/filter combination. These measurements were made in a manner similar to that mandated for MQSA, with the top surface of the ionization chamber of the exposure meter placed 4.5 cm above the tabletop, 4 cm from the chest-wall edge, with no backscattering material, and centered left to right in the image.

At a clinically relevant kV and the most commonly used target and filter combination, a series of exposures was made, starting at 20 mAs (for all systems except Fischer) and doubling the mAs setting repeatedly to analyze the linearity of the generator. Tube output (mR) was plotted against mAs, and a linear least-squares linear fit was applied. Four exposures were made with an mAs setting in the midclinical range to evaluate the short-term reproducibility of the generator, and again the COV among the measured outputs was calculated. The linearity and reproducibility exposures were

taken with the ionization chamber placed on top of a 4 cm thick phantom positioned approximately 4 cm anterior to the chest-wall edge of the detector.

b. Results and discussion. The output rates universally exceeded the MQSA required minimum of 7 mGy air kerma per second (102 measurements).

Most systems exhibited extremely good linearity of generator output with Pearson's correlation coefficients, r , describing the quality of a linear least-squares fit between measured mR and selected mAs close to 1. The minimum r value seen among all 136 testing instances was 0.996.

All systems exhibited very good short-term reproducibility of generator output with mean COVs less than the MQSA limit of 0.05. The maximum COV seen among all testing instances was 0.028, and the average COV value was 0.001.

6. Detector linearity and reproducibility

a. Method. The linearity and reproducibility of the detectors were tested using images of a 4 cm uniform slab of PMMA, obtained during the evaluation of tube output linearity and reproducibility. Before detector response could be analyzed, it was first related to the measured x-ray entrance exposure (in mR) including backscatter. For all systems except the Fuji system, a representative mean pixel value (MPV) was calculated by averaging over a 2×2 cm region of interest (ROI) in the unprocessed digital image. For the Fuji systems the "S-number," which reflects the *speed* of the entire plate, was recorded.

To analyze the linearity of the detector with respect to x-ray exposure, the MPV was plotted against entrance exposure, and a linear least-squares fit was applied, for all systems except the Fuji system. For the Fuji systems the constancy of the S-number multiplied by the exposure was evaluated by calculating the COVs for the series of images acquired over a range of mAs settings. In all cases, the mean value calculated from the four images (MPV or $S \times \text{mR}$), taken at a single mAs, was treated as a single measurement to avoid bias. Mean Pearson's correlation coefficients and COVs were determined over multiple units of each system and multiple testing sessions.

To analyze the short-term detector reproducibility, the COVs of MPV/mR and $S \times \text{mR}$ for four sequential images of the 4 cm thick uniform plastic slab, acquired with constant exposure factors, were calculated. Again, mean COVs were determined over all units for a system type and all testing sessions.

b. Results and discussion. The characteristics of linear least-squares fits to the signal versus entrance exposure (MPV vs mR) data for the most common target, filter, and kV combinations for each system type except Fuji are given in Table III. The Pearson's correlation coefficient data for all testing sessions are given in Table IV. The average COV for the $S \times \text{mR}$ values on the Fuji systems was 0.032, with a standard deviation of 0.079 (29 measurements). The maximum COV measured was 0.434, which was due to a faulty photomultiplier, and also resulted in an S number difference. In general, all systems exhibited excellent detector linearity.

TABLE III. Detector linearity: Characteristics of linear-least squares fits of signal versus entrance exposure for images of a uniform phantom. r is Pearson's correlation coefficient (N is the number of fits from separate inspections and σ is the standard deviation).

System	N	Target	Filter	kV	Mean r (σ)	Min r	Mean slope (σ)	Mean intercept (σ)
Fischer	18	W	Al	29	1.000 (0.001)	0.997	1.9 (0.3)	-8 (32)
GE	22	Mo	Mo	27	1.000 (0.000)	1.000	0.87 (0.05)	-6 (3)
Lorad DBI	5	Mo	Mo	26	1.000 (0.000)	1.000	1.0 (0.6)	-1 (4)
Lorad Selenia	5	Mo	Mo	27	1.000 (0.000)	1.000	0.21 (0.03)	43 (6)

All systems exhibited very good short-term reproducibility of detector response. The mean COVs for each system type were all less than 0.02, with standard deviations less than 0.02 as well. Out of 136 measurements, there was only one instance where the COV of a detector exceeded 0.05.

7. Half-value layer

a. Method. The half-value layer (HVL) was measured in a similar manner to that done for screen-film systems, at a single clinically relevant kV for each available target-filter combination. The HVL was required to be between the minimum and maximum values from the ACR Quality Control Manual² and Fischer's QC manual for the Senoscan. [SenoScan Full Field Digital Mammography System Operator Manual, Issue 1. Revision 2. Fischer Imaging Corporation, Denver (2001).]

b. Results and discussion. None of the 396 HVL measurements fell outside the recommended range. The stability of the HVL measurement was evaluated by calculating the COV among measurements made during different physics testing sessions using the same target, filter, and kV selections. The mean COVs for the systems were all less than 0.05. The average number of measurements used to calculate a single COV was 3. The HVL measurements showed little variation for most of the units.

8. Focal spot

a. Method. In this test, the effective spatial resolution in line pairs per millimeter (lp/mm) was measured on a mammographic screen-film receptor positioned on the breast sup-

port table. The test pattern was located 4.5 cm above the cassette, within 1 cm of the chest-wall edge of the breast support plate, with 2.5 mm of aluminum or 4 cm of acrylic between the pattern and the breast support table. This test was designed to evaluate the extent to which the focal spot size limits spatial resolution independent of the detector.

b. Results and discussion. All but five of 246 measurements passed the minimum MQSA requirement of 13 line pairs/mm in the direction parallel to the anode-cathode axis, and 11 line pairs/mm in the direction perpendicular to the anode-cathode axis. Of the five failing measurements, four were made on GE units and one was made on a Lorad DBI unit.

In all measurements made using the large focal spot, the measured resolution in this test was higher than the maximum theoretical resolution based on the del pitch in the digital image, i.e., the focal spot did not limit resolution. When using the magnification stand and small focal spot with the Fuji CR system, occasionally the effective resolution would fall below the maximum theoretical resolution based on the del pitch and the magnification factor. Under such circumstances, the focal spot could be the limiting factor for spatial resolution.

B. Image quality and radiation dose

1. Daily imaging of the accreditation phantom

a. Method. The mammography accreditation phantom² was imaged daily at each site by the technologist using a manual technique matching the exposure factors that would be selected by the automatic exposure system for the 4 cm uniform PMMA phantom. For the first 15 months of the trial, the images generated each day were collected and scored centrally by a standard reader. For the remainder of the trial, the phantom was imaged daily as a system check, but only one image per week was evaluated centrally.

b. Results and discussion. For this test, 5766 mammography accreditation phantom images were evaluated at the central physics site, by the same reference reader. Of these images, only 57 failed to meet the requirements of acceptable artifact levels and a minimum score of four fibers, three speck groups, and three masses as defined by the scoring system of the ACR Mammography Accreditation Program.

TABLE IV. Detector linearity: Pearson's correlation coefficients (r) for linear-least squares fits of signal versus entrance exposure for images of a uniform phantom (N is the number of fits from separate inspections and σ is the standard deviation).

System	N	Mean r (σ)	Min r
Fischer	31	1.000	0.997
GE	56	1.000	1.000
Lorad DBI	7	1.000	1.000
Lorad Selenia	13	1.000	0.996

TABLE V. Scores given by the reference reader for the accreditation phantom. N is the number of images, m is the mean score, σ is the standard deviation, and S indicates statistical significance ($p < 0.05$) for a t-test between the mean score for that system type and the maximum mean score.

System	N	Fibers				Speck groups				Masses			
		m (σ)	Max.	Min.	S	m (σ)	Max.	Min.	S	m (σ)	Max.	Min.	S
Fischer	1355	4.36 (0.3)	6	3.5	y	3.4 (0.4)	4.5	3	y	3.5 (0.3)	5	3	y
Fuji	1797	4.5 (0.3)	6	1.5	y	3.6 (0.4)	4	2	y	3.4 (0.3)	4.5	1	y
GE	2098	4.6 (0.3)	6	3.5	—	3.9 (0.3)	4.5	3	—	3.7 (0.4)	5	2.5	y
Lorad DBI	301	4.43 (0.2)	5.5	4	y	3.8 (0.3)	4	3	y	3.8 (0.3)	4.5	3.5	—
Lorad Selenia	189	4.3 (0.5)	5.5	0.5	y	3.5 (0.4)	4	2	y	3.8 (0.4)	4.5	3.5	n

This is a total failure rate of less than 1%. Among these 57 failing phantom images, four were acquired using obviously incorrect techniques, 23 were blank, and four had obvious artifacts, leaving only 26 failures where the phantom provided useful information (less than 0.45%). The mean phantom scores given by the reference reader for each machine type are shown in Table V. Images where the recorded score was 0 fibers, 0 speck groups, and 0 masses have been omitted. The differences in the mean phantom scores between machine types are less than 0.5 levels, but are statistically significant (t-test, $p < 0.05$), except for fibers and specks between the Fischer and Lorad Selenia systems, and for masses between the Lorad DBI and Lorad Selenia systems.

2. Weekly imaging of the uniform phantom

a. Method. The 4 cm thick uniform PMMA phantom was imaged once a week by the technologist, using automatic exposure control where available. On systems employing soft-copy display for image review, the mean (MPV) and standard deviation (Noise Index, or NI) of the pixel values in a region of interest in the center of the unprocessed image were recorded as a measure of signal level and noise. On systems employing hard-copy display (printed films), the optical density of a spot in the middle of the image was recorded. The variations in technique (mAs) and signal level were tracked to ensure that systems were stable. Note that the Fischer and Lorad Selenia systems did not have AEC, so a fixed mAs was used to image the phantom on those units.

TABLE VI. Number of days to which target values apply. σ is the standard deviation.

System	# of Target values	Mean (σ) Days	Min days	Max days
Fischer	34	109 (86)	7	329
Fuji	33	145 (101)	6	364
GE	36	225 (163)	18	650
Lorad DBI	8	124 (114)	21	379
Lorad Selenia	7	160 (81)	7	279

Sites calculated new target values whenever the unit was serviced or the imaging technique was changed.

b. Results and discussion. The mean number of target values calculated for each system type and the number of days for which those values were active are given in Table VI. This provides a rough indication of the stability of exposure techniques within the program for the digital systems. For each time period for which target values apply on each mammography unit, the COVs for the mAs, OD or MPV and NI were calculated.

AEC performance (mAs reproducibility) was generally excellent, with the mean COVs for all systems being less than 0.05. Signal levels and system noise in the images evaluated in soft copy were also very stable, with the mean COVs for the Fischer, GE, and Lorad Selenia systems being less than 0.05.

Summary information for the printed images can be found in Table VII. Measuring the printed optical density on films from the Lorad DBI system proved problematic. The Lorad DBI and Lorad Selenia systems automatically selected the look-up table used to print the image, so the user could not guarantee that the image was printed the same way each time, possibly invalidating optical density as a measure of system consistency. The automatic algorithm for determining printer parameters was not designed for the narrow distribution of pixel values in the image of the uniform phantom;

TABLE VII. Constancy of overall signal level: hard copy. Coefficient of variation of OD measured on printed films of weekly uniform phantom. Each COV was calculated for data where a single target value was in effect on a single mammography unit. Images of the uniform phantom were not printed on the Fischer and GE systems. σ is the standard deviation.

System	# of Target values	# of Data points		Optical density	
		Mean	(σ)	COV _{Mean}	COV _{Max}
Fuji	33	19	(14)	0.04	0.13
Lorad DBI	8	8	(5)	0.08	0.20
Lorad Selenia	2	11	(8)	0.03	0.06

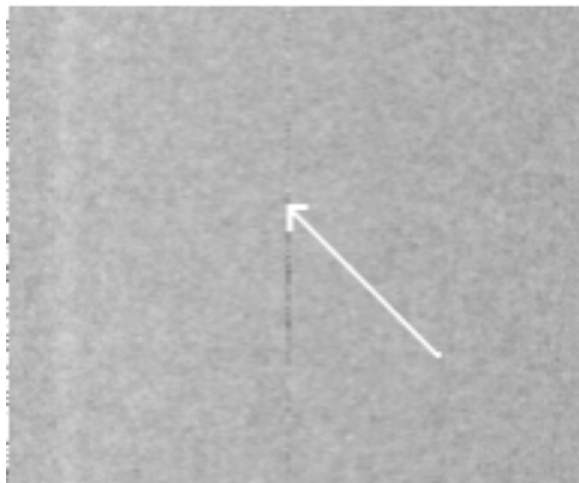


FIG. 6. Stitching artifact in a ROI from an image of a uniform phantom taken on a Fischer unit.

often the resulting films were too dark and had excessive contrast. This is reflected in the somewhat high maximum COV for optical density.

3. Artifacts

a. Method. A 4 cm thick uniform phantom covering the entire image field was imaged with all available focal spot and filter combinations. Artifacts were categorized according to their probable source, such as filter corrosion, grid-related, flat-fielding nonuniformities, and dead detector elements (dels).

b. Results and discussion. Several types of artifacts of varying severity were seen. Stitching artifacts were seen (Fig. 6) as subtle discontinuities at the “dead” gap between charge-coupled device (CCD) sections in the detector strip of the slot (Fischer) or the full-area detector (Lorad DBI). Because flat fielding of the x-ray field and detector response is not performed on the Fuji system, x-ray field nonuniformities (heel effect) were observed. Occasionally structural mottle (probably from the filter or grid, and not completely corrected by flat fielding) was seen on GE images (Fig. 7). Other artifacts included random pixels that were noticeably brighter or darker than the surrounding image (“salt and pepper,” caused by dead pixels and/or pixels with markedly dif-

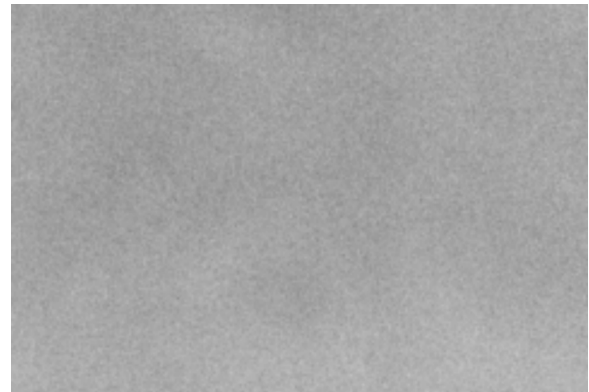


FIG. 7. Mottle artifact in a ROI from an image of a uniform phantom taken on a GE DMR 2000D unit. The mottle is visible diffuse regions of differing gray levels.

ferent gains) and other flat-fielding problems (different gains on different detector tiles) on Lorad DBI units.

4. Misty/Conspicuity test

a. Method. Images of the Misty phantom were taken using automatic exposure control or automatic technique selection (AEC) or an appropriate manual technique for 5.5 cm of PMMA. For the purposes of this document, except where noted otherwise, the term AEC refers both to systems which control only exposure time and to those which control multiple exposure factors such as kV, target, and filter. Where possible, images were evaluated on the soft-copy display. In the Misty phantom, the star conspicuity pattern is replicated over two thicknesses of PMMA attenuator: 5.4 and 2.8 cm. Display windows and levels were set by the viewer to best display the star groups over the thickness of material being scored. The viewer was allowed to apply a zoom function as desired.

At the time of the study, the Fuji and Lorad DBI systems and some Lorad Selenia systems did not have soft-copy display capability. Where possible, two versions of the image were printed, each adjusted to display the stars optimally over one thickness region. Where printing was not available or the algorithm could not accommodate the DMIST phan-

TABLE VIII. The mean number of groups of stars resolved as specks and the mean number of groups of stars in which the points could be resolved over 5.4 and 2.8 cm of PMMA for each system (N is the number of measurements from separate inspections, σ is the standard deviation among scores, and S is whether there is a statistically significant difference from the maximum mean score, t-test, $p < 0.05$).

System	N	5.4 cm				2.8 cm			
		Specks (σ)	S	Points (σ)	S	Specks (σ)	S	Points (σ)	S
Fischer	30	12.9 (1.0)	...	3.0 (1.3)	...	13.8 (1.0)	n	4.8 (1.4)	n
Fuji	27	11.5 (1.2)	y	2.3 (0.9)	y	13.0 (2.8)	n	3.6 (1.4)	y
GE	54	11.7 (1.6)	y	2.7 (2.8)	n	13.1 (2.1)	y	4.1 (2.4)	n
Lorad DBI	7	12.3 (2.1)	n	2.1 (1.3)	n	12.1 (4.9)	n	5.4 (0.8)	...
Lorad Selenia	12	12.3 (1.8)	n	2.8 (1.6)	n	13.8 (0.4)	...	4.8 (1.0)	n

TABLE IX. Characteristics of linear-least squares fits of variance versus signal for images of a uniform phantom. r is Pearson's correlation coefficient (N is the number of fits from separate inspections and σ is the standard deviation).

System	N	Mean r (σ)	Min r	Max r	Mean slope	Mean intercept
Fischer	31	0.993 (0.005)	0.978	1.000	0.16	-22
Fuji	29	0.965 (0.060)	0.803	1.000	0.36	24
GE	56	0.998 (0.002)	0.990	1.000	0.17	-24
Lorad DBI	7	0.985 (0.012)	0.965	0.996	0.45	-154
Lorad Selenia	13	0.994 (0.005)	0.982	1.000	0.15	-8

tom (the Lorad DBI and some Lorad Selenia systems), the images were scored on an independent image display workstation.

The groups of stars of decreasing size in the Misty phantom were scored in two ways: (1) The smallest group where all six stars were visible was recorded. (2) One of the stars in each size group is missing a point. If the star with the missing point could be identified, that group was counted as a measure of the ability to visualize detail. In both cases, counting ceased at the first group that could not be discerned.

b. Results and discussion. The results of the conspicuity test using the Misty phantom, as reported by the site physicists, are given in Table VIII. In the third column of the table the mean value of the number of the group (5.4 cm of PMMA) containing the smallest specks that are detectable with each system is given. The variability in this measure among the different machines of each type and different test visits is shown in parentheses as a standard deviation. In the fourth column the groups and standard deviations over 5.4 cm of PMMA are reported for identification of the missing point. On the right side of the table, similar results are provided for the 2.8 cm PMMA attenuator.

As expected, the conspicuity of the groups of stars is higher where the attenuating PMMA is thinner, as there is less x-ray noise, due to a greater number of detected direct x-ray photons, as well as fewer scattered photons. The ability to detect the stars as specks and to resolve the points of the stars does not correlate with the nominal detector element (del) size. The 100 μm GE system has comparable scores to

the Fuji 50 μm CR system as well as the Lorad DBI (40 μm del) system (p value >0.05). Statistically significant differences were found between the Fischer and Fuji systems, which have the same del size.

5. Noise levels and noise power spectrum

a. Method. To approximate a clinical situation for noise evaluation, images of a 4.0 cm thick, large-area uniform slab of PMMA were obtained at clinically relevant target, filter, and kV combinations. An ionization chamber was positioned on the upper surface of the slab to track relative changes in entrance exposure. A series of images was taken, doubling the mAs between exposures.

Noise is quantified as the variance of pixel values in a ROI, 4.0 cm^2 in area. Variance was plotted versus signal level (MPV). For quantum-noise-dominated operation, a linear relationship was expected. For the Fuji CR system, which applies a logarithmic transform to the image data, variance was plotted versus the S-Number. If the system is quantum noise limited, these should be linearly related as both S and the variance of the signal are proportional to $1/E$, where E is exposure. The signal is equal to $c \times \log(E)$, where c is a constant. Its variance (neglecting other sources of noise), will therefore be $[\partial c \log(E) / \partial E]^2 \sigma_E^2$, which simplifies to $c^2(1/E^2)E$ or c^2/E .

TABLE X. Quantum vs nonrandom noise—The ratio between the standard deviation among pixel values of a 20×20 mm ROI, in an image formed by averaging four images, and the standard deviation among pixel values of the same ROI in a single image, averaged for all units of each system. The standard deviation (σ) is shown for the values averaged among each unit type. If all noise were due to Poisson fluctuation, the ratio would be 0.5.

System	N	Noise 4 avg/single (σ)	Noise _{Max} 4 avg/single
Fischer	27	0.56 (0.04)	0.66
Fuji	24	0.70 (0.08)	0.95
GE	54	0.54 (0.07)	0.68
Lorad DBI	6	0.56 (0.03)	0.61
Lorad Selenia	13	0.54 (0.04)	0.58

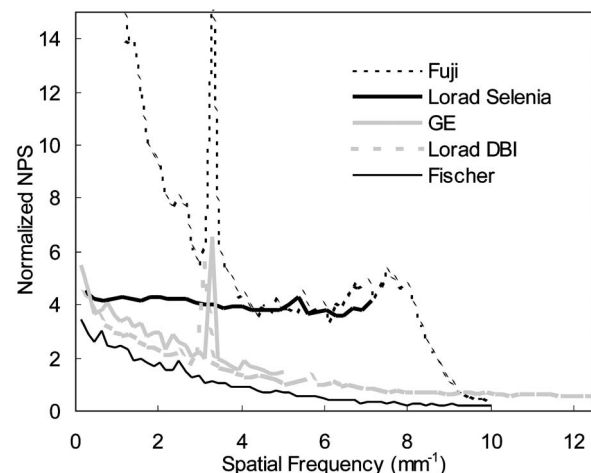


FIG. 8. Typical noise power spectra for each system.

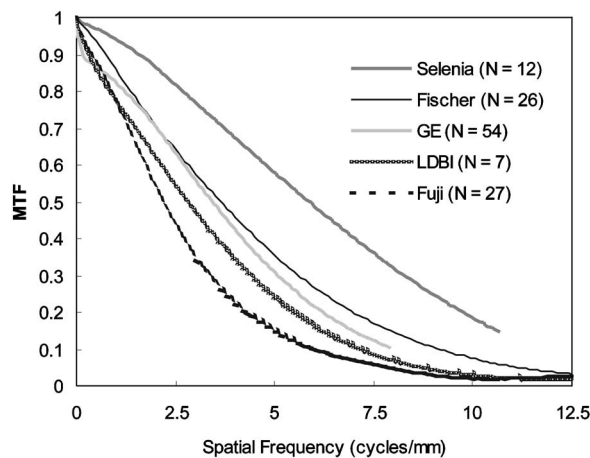


FIG. 9. Average presampled modulation transfer functions for the different systems tested. Images were taken with the tool 4.5 cm above the tabletop.

The level of nonrandom noise in the unprocessed images was estimated by comparing the standard deviations between pixels in different 4.0 cm^2 areas in a single image against the standard deviations in those areas, in the average of four images taken at the same technique. When quantum noise is the only noise source, random noise scales by $1/\sqrt{N}$, where N is the number of images averaged. Hence, one would expect the standard deviation in the average of four images to be half that of a single image. Any increase in standard deviation provides an estimate of the level of nonrandom noise.

Noise power was measured from four exposures of the PMMA slab with identical machine settings (clinically appropriate technique factors for a 4.5 cm thick breast composed of 50% fibro-glandular tissue and 50% fatty tissue). The NPS was calculated on each unprocessed image using the algorithm outlined by Dainty and Shaw,¹⁷ dividing a $7 \times 7 \text{ cm}$ central region into squares, each with 64 or 128 pixels on a side. The zero-frequency offset and low-frequency background trends were removed by subtracting a low-pass filtered version of the image.¹² The resulting spectra were compared to the ideal case by dividing by $\text{NPS}_{\text{ideal}} = k^2/\phi$, where k is the average pixel value in the central region and ϕ is the estimated fluence. To estimate fluence, spectra were measured on GE, Selenia, and Fischer systems. We assumed that the spectra for the Lorad DBI and Selenia

generators were the same. For the Fuji system, we used the GE Mo/Mo spectra. Fluence was scaled by mAs. The four estimated noise power spectra for each system were then averaged together to reduce uncertainty in the estimate.

b. Results and discussion. Pearson's correlation coefficient, r , describing the quality of a linear least-squares fit between signal level and noise are shown for the systems in Table IX. The coefficients were averaged over all units for each system.

The ratios of the standard deviation for a ROI in a fixed location, averaged over four images, to the value for a single image are shown in Table X. The ratios shown were obtained by averaging over all testing periods for a given system type. Finally, typical normalized noise power spectra for each system, in the direction parallel to the chest-wall edge of the detector, are shown in Fig. 8. It should be noted that the units of the spectra shown here have been normalized using estimated fluence, and therefore, inferences should not be made based on their relative magnitudes. The spikes noted in the NPS for the Fuji, GE, and Lorad DBI systems are most likely caused by very slight hesitations in the motion of the grid, as the frequency matches that of the typical spacing used in grids for mammography.¹⁸ The standard error in the NPS estimates is approximately 5%.¹⁹

All systems showed a strong linear relationship between variance and signal, indicating that, for most systems, the noise is close to being quantum limited when measured under these experimental conditions.

As seen from Table X, the Fuji system has the highest amount of nonrandom noise. This is to be expected, as no flat fielding is employed to remove the influence of heel effect and tabletop structure and variation of sensitivity both within and between image plates. Note that this measure of nonrandom noise does not take into account the pixel-to-pixel correlation present on some systems due to blurring in the phosphor. These results indicate that in all of the digital systems, except the CR system, the effects of structural noise have been almost entirely removed by the flat-fielding operation.

6. Effective system modulation transfer function

a. Method. The MTF tool was taped to the underside of the compression paddle at a slight angle (approximately 1:17

TABLE XI. Percentages of the Nyquist frequency for each system at which the presampled modulation transfer functions drop to 50%, 10%, and 4% (N is the number of measurements from separate inspections).

System	N	50%		10%		4%	
		Mean %	Min %	Mean %	Min %	Mean %	Min %
Fischer	32	40	24	98	48	123	66
Fuji	28	22	19	60	43	85	60
GE	54	65	52	>149	112	>149	127
Lorad DBI	7	22	20	57	50	77	64
Lorad Selenia	12	82	56	>149	129	>149	145

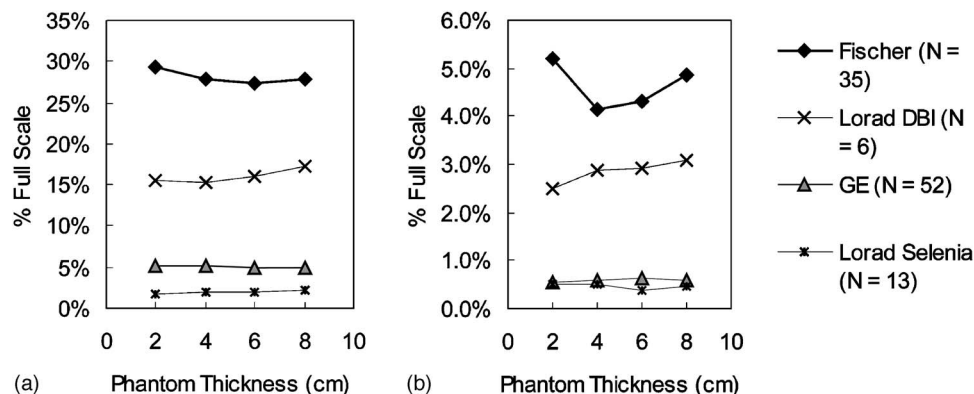


FIG. 10. (a) MPV as percentage of full scale vs phantom thickness; (b) NI as a percentage of full-scale vs phantom thickness.

slope was requested) with respect to the sides of the paddle (Fig. 3). The paddle was set 4.5 cm above the tabletop. Clinically used kV, target, and filter combinations were employed, with mAs selected such that no pixels within the image of the tool had either the maximum (saturated) or minimum (no signal) possible values. Software was written to calculate the presampled MTF (spatial frequencies referred to the detector plane) for each slanted edge¹² using the method described by Fujita *et al.*²⁰ The results for all units of a particular system type were averaged together to obtain the MTF for that system.

b. Results and discussion. The presampled MTFs measured on several systems are shown in Fig. 9 and Table XI. The GE and Lorad Selenia systems have MTFs that stay relatively high ($>10\%$) beyond the Nyquist sampling limit of the detector. This can result in aliasing of higher-frequency information. At a spatial frequency of 2 mm^{-1} , the standard deviation of the MTF values was between 0.02 and 0.05 for each system type.

The fact that the relative ordering of the MTFs of the different systems does not reflect their inverse del sizes demonstrates that the del dimensions alone do not fully predict system spatial resolution. The system with the highest MTF (the Lorad Selenia) does not have the smallest del size ($70 \mu\text{m}$). It is, however, the only system that directly converts the energy of interacting x rays to an electronic signal, rather than converting them to light in a phosphor as an intermediate step. The CsI phosphor-based systems (Fischer, GE, and Lorad DBI) all have quite similar MTFs, suggesting that light scatter in the phosphor may be limiting the resolu-

tion. The system with the lowest MTF, Fuji, has one of the smallest del sizes ($50 \mu\text{m}$). The lower MTF may be due to scatter of laser light in the phosphor during the read-out process, or imprecision in the motion of the laser beam or the plate in the reader. It should be noted that the clinical images produced by the Fuji system are logarithmically rescaled and processed with proprietary signal processing algorithms. Because the MTF calculations were carried out on raw images,

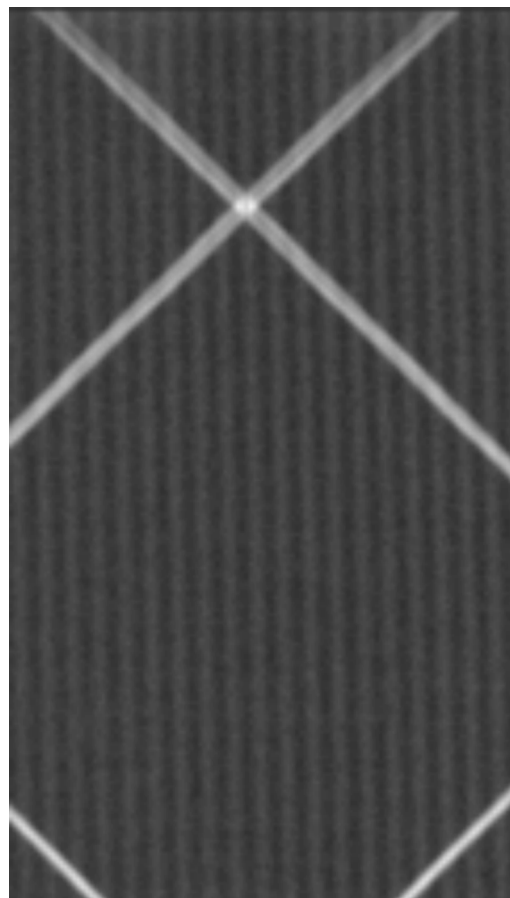


FIG. 11. Region of interest from an image taken of the distortion test tool on a Fischer unit, showing blurring at the start of image acquisition (top of image) when the mechanical scanning speed was not synchronized with data read-out. As the gantry motion comes up to speed, the image gets sharper, which is why the lines near the bottom of the image are in focus. This distortion can be minimized with proper system configuration.

TABLE XII. Inverse sensitivity numbers (S^{-1}) vs phantom thicknesses on the Fuji CR system. N is the number measurements from separate inspections and σ is the standard deviation.

Sensitivity—Fuji system $N=19$		
Thickness cm	Mean	(σ)
2	0.011	(0.0045)
4	0.012	(0.0046)
6	0.013	(0.0046)
8	0.013	(0.0045)

TABLE XIII. Mean signal-to-noise ratio (SNR), defined as MPV/NI for different thickness and systems (N is the number of measurements from separate inspections and σ is the standard deviation).

System	N	Thickness (cm)			
		2	4	6	8
		mean (σ)	mean (σ)	mean (σ)	mean (σ)
Fischer	35	103 (14)	97 (15)	89 (11)	84 (12)
GE	52	78 (8)	73 (8)	67 (7)	66 (7)
Lorad DBI	5	70 (10)	68 (10)	64 (9)	58 (12)
Lorad Selenia	13	49 (7)	48 (7)	44 (5)	45 (3)

in linear (x-ray fluence) space, the effect of the processing is not seen in the MTF. Clearly, such algorithms may enhance the MTF, the limit of such processing being the risk of amplifying the appearance of image noise as well. The acceptable degree of processing is, therefore, related to the inherent relationship between signal and noise in the image data.

7. Thickness tracking

a. Method. Thickness tracking was evaluated by imaging 2, 4, 6, and 8 cm thick slabs, composed of either PMMA or BR-12, a breast tissue-mimicking plastic using the AEC. Where no AEC was available, exposure factors were chosen from the clinical technique chart. The signal-to-noise ratio (SNR) for a region of interest was calculated for each unprocessed image as $\text{SNR} = \text{MPV}/\text{NI}$. The standard deviation among MPVs for each system type at each phantom thickness was also calculated.

The SNR was not examined on the Fuji system, because the exact relationship between the pixel value and the transmitted radiation exposure was not known. The information required for image rescaling was not available in the image header. Instead, for this system the sensitivity number (S-Number) provided by the plate reader, which is inversely proportional to the amount of radiation detected by the plate, for a constant beam quality, was examined.

b. Results and discussion. The MPV and NI in a central ROI approximately 4 cm² in area produced by each system for different phantom thicknesses are shown in Fig. 10. In each case the tests were conducted with the system operated under typical clinical conditions. Values are expressed as a

percentage of the full scale provided by each system to give a sense of how the available dynamic range of the detector is being used. The full-scale values are Fischer—4095; Fuji—1023; GE, Lorad DBI, and Lorad Selenia—16383. These data are given for all systems except the Fuji system, for which the inverse of the mean sensitivity number (S-Number) for different phantom thicknesses is given in Table XII. The mean SNRs (over all units and test sessions) are given in Table XIII.

From Fig. 10(a) and Table XII, it appears that the AEC or manual technique charts are being used primarily to maintain a constant mean pixel value (rather than SNR) across the range of breast thicknesses and compositions. Also notable is the wide variation among manufacturers in the choice of the targeted pixel value as a percent of the dynamic range of the detector (between ~30% and ~2%). Note that the extremes in this regard are the Fischer Senoscan and the Lorad Selenia, which do not have automatic exposure control. As thickness increases, the SNRs appear to drop for the Fischer, GE, and Lorad DBI systems.

8. Geometric distortion

a. Method. The distortion test tool was imaged using a manual technique selected to provide good contrast between the circuit board substrate and the copper lines. The resulting images were examined for any blurring, bending of the lines, or discontinuities.

b. Results and discussion. Occasional blurring was noted in images acquired on the Fischer system at the beginning and end of the scan. The blurring at the beginning of the scan

TABLE XIV. Mean glandular dose for the standard breast (N is measurements from separate inspections).

System	N	Entrance exposure		Range mR	MGD mGy	Range mGy
		$\mu\text{C}/\text{kg}$	mR			
Fischer	26	144	560	400–750	1.31	0.81–1.79
Fuji	25	278	1080	270–1660	1.87	0.79–2.69
GE	46	196	760	410–1410	1.49	0.84–2.53
Lorad DBI	11	294	1140	540–1610	1.98	1.10–2.70
Lorad Selenia	13	273	1060	620–1830	1.85	1.09–2.98
All	121	219	850	270–1830	1.62	0.79–2.98
Screen-film	149	304	1178	708–1810	1.90	1.24–2.72

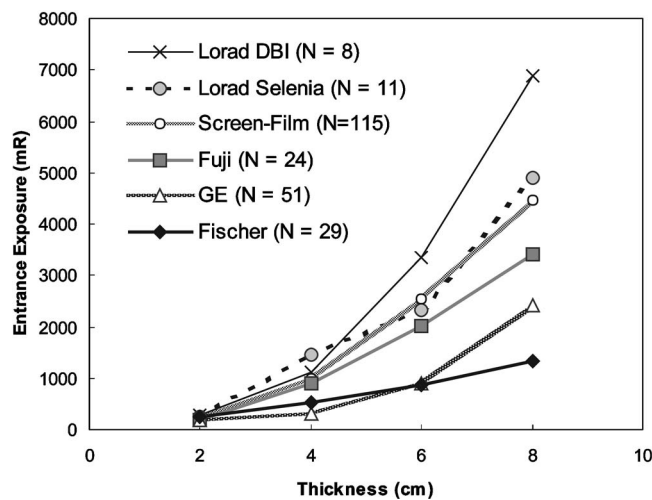


FIG. 12. Estimated entrance exposures for 2, 4, 6, and 8 cm phantoms expressed in mR. N is the number of measurements from separate inspections.

indicates that data collection was occurring before the correct scan speed was reached, when the motion of the detector would not yet be matched to the integration and readout rate. Similarly, blurring at the end of the scan suggests data collection was taking place while the detector assembly was decelerating at the end of the scan (Fig. 11). In one instance, blurring was noted in a band across the center of the image. In that case, it was discovered that an obstruction was causing a decrease in the speed of the detector assembly, disrupting the synchronization with the data integration and readout. Dead gaps between tiled detector elements on the Fischer and Lorad DBI systems were evident as subtle discontinuities in the copper lines.

9. Entrance exposure and mean glandular dose

a. Method. Images of 4 cm of PMMA and 2, 4, 6, and 8 cm of BR12 or PMMA were obtained. Where it was available, the AEC selected the exposure parameters; otherwise, the posted technique chart was followed. The dose estimation is that which would be reported to an accreditation body, by a medical physicist, and is based on the standard clinical practice of the site. Entrance exposures were estimated from the technical factors recorded and the tube output measurements (mR/mAs) made during the physics survey. SFM

TABLE XV. Failure rates (%) in review workstation monitor evaluations.

N (number of tests)	Fischer 31	GE 60	Selenia 7
Distinct gray levels—Fails	3	2	14
Line pairs visible—Fails	23	45	29
0%–5% contrast fails	3	2	14
95%–100% contrast fails	6	2	0
Streaking/smearing present	26	5	0

doses were taken from the SFM unit physics surveys done by the local physicist. Based on the measured HVL of the system at the selected kV, the entrance exposure (mR) for 4 cm of PMMA (approximately equivalent to 4.2 cm 50% fat–50% fibro-glandular tissue) was converted to mean glandular dose (MGD) (mGy) using published conversion factors^{21,22} for all systems except the Fischer system, where values supplied by the manufacturer were used [SenoScan Full Field Digital Mammography System Operator Manual, Issue 1. Revision 2. Fischer Imaging Corporation, Denver (2001)].

b. Results and discussion. The estimated MGD values are given in Table XIV. The average estimated entrance exposures for different breast thicknesses are shown in Fig. 12.

The MGDs on all systems except the Fischer Senoscan were similar to screen-film units (1.5–2.0 mGy). This system had the lowest entrance exposure and mean glandular dose because there is no grid and it employs a more penetrating beam (W anode, Al filter, higher kVs) than the other systems.

A broad range of exposures and doses was found with the Fuji system because the type of conventional mammography unit(s) used with the CR plates varied from site to site. Much of the variation in entrance exposures and doses on all systems is attributable to the different choices made in technique selection at the sites.

The Lorad DBI system had the highest entrance exposures for the 4, 6, and 8 cm phantoms because some sites with this system chose not to use the rhodium filter. The variation in filter use also accounts for the high variation in exposures between sites for this system.

The Fuji and Lorad DBI systems had the highest MGDs. For Fischer and GE systems, exposures increased less rapidly with increasing breast thickness than for the other digital systems (Fig. 12). This is due to the more penetrating beams

TABLE XVI. Different measures of how well monitor calibration matches the DICOM gray-scale display function. N is the number of test measurements for each system type. σ is the standard deviation. $|r|$ is the absolute value of Pearson's correlation coefficient and SE is the estimated standard error for the fit.

System	N	LUM		Linear fit coefficients		$ r $		SE	
		Mean	Min	Mean _{Slope} (σ)	Mean _{Intercept} (σ)	Mean	Max	Mean	Max
Fischer	62	0.39	0.11	0.01 (0.07)	2.04 (0.48)	0.45	0.98	0.34	0.83
GE	118	0.45	0.12	−0.001 (0.10)	2.08 (0.73)	0.39	0.97	0.43	4.49
Lorad		1.57	0.95						
Selenia	14			0.44 (0.24)	−0.13(1.26)	0.88	0.99	0.50	1.19

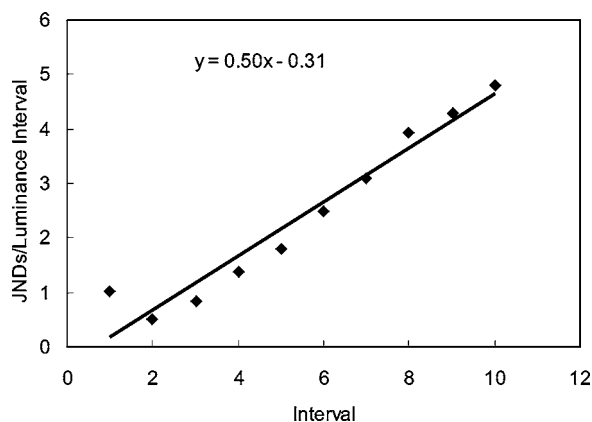


FIG. 13. Typical JNDs/luminance interval for a monitor from a Lorad Selenia image display workstation. Since the workstation is not calibrated to the DICOM standard, a linear fit to the data does not result in a horizontal line.

used with these systems for thicker breasts. The GE system switches to Rh target/ Rh filtration, and higher kVs at lower thicknesses than other systems. This suggests that there may be a dose reduction compared to screen-film units for women with larger breasts with these systems.

C. Display

1. Monitor evaluation

a. Overall display quality—SMPTE pattern.

i. *Method.* The SMPTE pattern was displayed on all the workstation monitors. Different elements of the SMPTE pattern were examined by the site physicist to evaluate display quality. The gray-level patches surrounding the center of the pattern provided a qualitative sense of monitor luminance response. They should be distinguishable from one another. The 0%–5% and 95%–100% contrast patches were inspected to verify that the inner and outer squares were distinguishable, such that contrast information in the darkest and brightest ranges was maintained. The line-pair patterns were examined to check display resolution. It was expected that the finest line-pair patterns in all locations in the image should be resolved. The image was inspected for evidence of streaking, smearing, or other display-related artifacts.

ii. *Results and discussion.* The failure rates for the different elements of the SMPTE pattern are given in Table XV.

The failures in the line-pair tests were mainly in the horizontal direction, due to the limiting bandwidth of the video amplifier.²³ The physicist had the discretion to pass the monitor even if it failed a subtest.

b. Monitor luminance response measurement.

i. *Method.* The luminance response of the monitors attached to the radiologist's reading station was measured by the physicist for compliance to the DICOM Gray Scale Display Function (GSDF).^{24,25} The GSDF uses the Barten model for the response of the eye^{26,27} to ensure consistent image presentation across different hardware platforms by distributing the available gray levels in a perceptually linear fashion across the dynamic range of the display.

The protocol described by the DICOM Working Group for this measurement²⁴ was followed except that no correction was performed for the presence of ambient light falling on the monitor screen and that the gray-level patches on the SMPTE pattern were used instead of the larger area standard target. The number of just-noticeable differences (JNDs) between gray levels was calculated to test whether the monitor calibration adhered to the function across the full dynamic range of the monitor. The standard deviation in this quantity is referred to as the "LUM" measurement in the DICOM standard.²⁴ A linear least-squares fit of the JNDs per luminance interval versus luminance interval was performed for each monitor and each QC testing session.

ii. *Results and discussion.* The ensemble averages and minima of the LUM values for each system type are given in the first part of Table XVI. The mean and standard deviation of the slopes, and the y intercepts for the linear fits of each system type are also given in the table. The smaller the value of LUM and the closer the slope of the linear fit is to zero the closer the monitor's luminance response is to the GSDF. The monitors used in the Lorad Selenia units demonstrated the best linear fit, but the line is not horizontal and the LUM values are higher.

For Fischer and GE systems, adherence to the GSDF was deemed acceptable when the monitors were calibrated according to manufacturer's specifications, as shown by the small minimum LUM value and small slope; acceptance criteria are currently under review. The soft-copy display station for the Lorad Selenia was not designed to be calibrated

TABLE XVII. Different measures of how well laser printer calibration matches the DICOM gray-scale display function, assuming a viewbox brightness of 4000 cd/m² and no ambient light. " $|r|$ " is the absolute value of Pearson's correlation coefficient. SE is the estimated standard error of the fit.

Printer type	N	LUM		Linear fit coefficients		$ r $		SE	
		Mean	Min	Mean _{Slope} (σ)	Mean _{Intercept} (σ)	Mean	Max	Mean	Max
Agfa 5200	9	0.062	0.026	0.0014 (0.001)	0.20 (0.08)	0.52	0.94	0.05	0.09
Fuji DPL	15	0.103	0.066	−0.0134 (0.010)	0.30 (0.05)	0.52	0.81	0.09	0.15
Kodak 8610	18	0.111	0.077	−0.0290 (0.012)	0.37 (0.07)	0.77	0.92	0.07	0.09

TABLE XVIII. Artifact detection rates for the laser printers used in DMIST. N is the number of tests.

Printer	N	Artifacts	Rate (%)
Agfa LR5200	13	4	31
Agfa 4500	1	0	0
Fuji FM-DPL	13	7	54
Kodak 8610	27	7	26

to the DICOM standard, and consistently produced an uneven distribution of JNDs. An example of this behavior is given in Fig. 13.

Although our techniques were applied consistently, our use of small squares for the luminance measurement and failure to correct for ambient light may have limited accuracy of this test.

2. Laser printer evaluation

a. Printer calibration and artifacts.

i. Method. Hard-copy display systems were evaluated in a similar manner to that used for the monitors. At the outset of the study, we did not impose specific requirements for the calibration of the laser printer and systems were calibrated by the service engineer according to each vendor's specifications. As we gained experience in the program, we began to evaluate how well the calibration of the laser printer matched the DICOM GSDF. An image of the SMPTE test pattern was printed from the review workstation, where possible, by the physicist at equipment evaluations. Otherwise, a stored pattern was printed from the printer. The optical densities for the gray patches in the pattern representing various percentages of the full-scale signal were measured. To evaluate conformance to the GSDF, the optical densities were first converted to light output assuming a viewbox luminance of 4000 cd/m² and no ambient room light, and then the JNDs were calculated. A uniform image was printed to assess printer artifacts.

ii. Results and discussion. Results for the printer calibration evaluation are given in Table XVII.

The mean and minimum standard deviation (LUM) in the number of JNDs between 10% steps on the SMPTE pattern

or calibration gray-scale step wedge (whichever could be printed on a particular system) are given. The results presented have been averaged over all printers and test sessions for each system type. Again, the smaller the standard deviation, the better the conformation of printer response is to the standard.

A summary of the artifact detection rates is given in Table XVIII. These were mostly minor nonuniformities. Excessive artifacts were noted on only two of the images. These were corrected by servicing the printer.

b. Printed MAP image.

i. Method. As part of the physicist's evaluation of the laser printer, an image acquired of the MAP was printed and evaluated subjectively following the ACR phantom scoring procedure.² The mean scores for each test object and printer type were tested for statistically significant differences from the highest of the mean scores using the t-test and a p value <0.05 . For the Agfa 4500, which had only one image, the score was deemed significantly different if it fell outside of the 95% confidence interval of the highest mean score.

ii. Results and discussion. The mean scores the physicists assigned to the printed images of the MAP for each printer type are given in Table XIX. All printed images met the minimum required score of 4 fibers, 3 speck groups, and 3 masses.

c. Printer sensitometry.

i. Method. Sensitometry was tracked on the printers every operational day in a manner similar to that used for SFM. Where possible, an SMPTE pattern was printed (12 sites) and optical density measured in selected squares. At the other sites, the printer's built-in density test pattern was used (12 sites). Base+fog ($B+F$), mid-density (MD) (50% square), contrast or density difference (DD) (20%–80% squares) and maximum optical density (DMax) were charted. Those sites using the built-in test pattern selected steps in a manner similar to that used for screen-film sensitometry; selecting the step closest to the working optical density for the mid-density measurement, and measured 2 steps above and 2 steps below the mid-density step for the contrast measurement.

ii. Results and discussion. Average sensitometry values measured on the different printers are given in Table XX.

TABLE XIX. Scores given by the physicists to printed images of the accreditation phantom. N is the number of images, m is the mean score, σ is the standard deviation, and S indicates statistical significance ($p < 0.05$) for a t-test between the mean score for that system type and the maximum mean score.

Printer	N	Fibers				Speck groups				Masses			
		m (σ)	Max.	Min.	S	m (σ)	Max.	Min.	S	m (σ)	Max.	Min.	S
Agfa 4500	1	5.0			n	4.0			y	4.0			y
Agfa LR 5200	15	4.6 (0.5)	5.0	4.0	y	3.4 (0.4)	4.0	3.0	n	4.3 (0.4)	5.0	4.0	...
Fuji FM DPL	21	5.0 (0.6)	5.5	4.0	n	3.6 (0.4)	4.0	3.0	...	4.0 (0.4)	4.5	3.0	n
Kodak 8610	30	5.1 (0.6)	6.0	4.0	...	3.4 (0.3)	4.0	3.0	n	3.8 (0.5)	5.0	3.0	n

TABLE XX. Sensitometry values measured on the printer test pattern averaged over the N measurement points for each system type. σ is the standard deviation.

	N	B+F or LD	MD	DD	DMax
		OD (σ)	OD (σ)	OD (σ)	OD (σ)
Agfa 4500 ^a	105	0.29 (0.04)	1.49 (0.08)	1.08 (0.07)	3.50 (0.12)
Agfa LR5200 ^a wet	1027	0.19 (0.01)	1.57 (0.26)	1.73 (0.32)	3.39 (0.23)
Fuji FM-DPL	2332	0.19 (0.02)	1.31 (0.13)	1.64 (0.12)	3.49 (0.73)
Kodak 8610	1259	0.20 (0.01)	1.33 (0.10)	1.67 (0.10)	3.54 (0.07)

^aThe sensitometry on the Agfa 4500 and the Agfa LR5200 was measured on the SMPTE test pattern.

Values were averaged on the measurements from built-in test patterns for the Fuji FM-DPL and Kodak 8610 and from the SMPTE pattern for the Agfa 4500 and Agfa LR5200. These values are averaged across all sites with that printer model using the same test pattern and all dates that sensitometry data were collected.

Sites calculated new target values whenever the printer was serviced. Coefficients of variation were calculated among sensitometry values for each time period to which a target value applied to determine the stability of the printers. The COVs are given in Table XXI. These tables indicate that the sensitometry for the printers was quite stable, with most mean COVs below 0.05.

3. Soft-copy viewing conditions

a. Method. To help ensure that the ambient light levels were not degrading the quality of the soft-copy images at the radiologist's review workstation, the physicist assessed the viewing conditions in the soft-copy reading area. Photometric measurements were made on the face of the monitors under typical viewing conditions to measure diffuse incident light.

b. Results and discussion. All facilities met the DMIST requirements of 10 lux maximum incident illuminance at the faces of the monitors.

V. CONCLUSIONS

The QC tests demonstrated that the digital mammography equipment used in DMIST behaved in a consistent and repeatable manner for all machine types. The QC tests were useful in determining that the systems were operating correctly; however, they did not predict equipment problems (mechanical, detector, and computer) as these occurred without warning. The measurements presented here represent the state of equipment in the field during the image acquisition phase of the DMIST trial. In many cases, the systems have since been refined, and, therefore, performance is likely to have changed to some extent.

Because the user and display software can compensate for changes in equipment performance, subjective assessment of severity of artifacts, variability of system signal levels and noise characteristics are unreliable. For example, the failure rate on the MAP was extremely low. Similarly, gradual degradation of spatial resolution is difficult to detect subjectively. This motivates the use of quantitative tests like MTF and signal difference to noise ratio (SDNR), which are easy to perform on a digital system and are not observer dependent. If subjective assessment of image quality must be done, it is especially important to establish a consistent image viewing protocol for the tests.

Because of the relatively short duration of this study, further investigation of the long-term failure characteristics of

TABLE XXI. Constancy of printer performance. Coefficients of variation for sensitometry measures made on the printer test patterns. N is the number of different target values.

Printer	N	B+F COV		MD COV		DD COV		Max OD COV	
		Mean (σ)	Max	Mean (σ)	Max	Mean (σ)	Max	Mean (σ)	Max
Agfa 4500 ^a	4	0.10 (0.05)	0.14	0.04 (0.01)	0.06	0.06 (0.02)	0.07	0.03 (0.02)	0.06
Agfa LR5200 ^a	17	0.03 (0.01)	0.05	0.04 (0.01)	0.06	0.04 (0.04)	0.17	0.02 (0.01)	0.04
Fuji FM-DPL	27	0.04 (0.02)	0.12	0.03 (0.03)	0.13	0.03 (0.02)	0.08	0.02 (0.02)	0.08
Kodak 8610	16	0.04 (0.01)	0.06	0.02 (0.01)	0.05	0.02 (0.01)	0.05	0.02 (0.04)	0.16

^aThe sensitometry on the Agfa 4500 and the Agfa LR5200 was measured on the SMPTE test pattern.

digital mammography systems is advisable. Systems with mechanical scanning components in the acquisition or read-out stages may require more frequent monitoring because of their susceptibility to mechanical misalignment or temporal drift. Performance of the display system is critical and must be evaluated rigorously and at regular intervals.

Finally, with digital mammography there is the opportunity through automation of many of the routine QC tests to improve efficiency, to reduce variability and labor costs, and to allow technologists to focus on imaging patients. Based on our experience in DMIST, we believe that it is reasonable to streamline QC for digital mammography considerably and this is discussed in more detail in Part II.

ACKNOWLEDGMENTS

The authors gratefully acknowledge funding from the American College of Radiology Imaging Network and the National Cancer Institute, and the collaboration of the local physicists and technologists at all sites involved in the Digital Mammography Imaging Screening Trial. Part of the development of QC tests was supported by the Ontario Research and Development Challenge Fund. Betsy Joseph, MSc. was responsible for scoring all of the accreditation phantom images.

- ¹E. Pisano, R. E. Hendrick, S. Masood, and C. Gatsonis, *American College of Radiology Imaging Network—ACRIN 6652—Digital vs. Screen-Film Mammography*, ACRIN (2001).
- ²R. E. Hendrick *et al.*, *American College of Radiology (ACR) Mammography Quality Control Manual for Radiologists, Medical Physicists and Technologists* (American College of Radiology, Reston, VA, 1999).
- ³MSQA. Mammography Quality Standards Act Regulations 21 CFR PART 900—Mammography (2002).
- ⁴G. E. Mawdsley, M. J. Yaffe, D. A. Maidment, L. Niklason, M. B. Williams, and B. M. Hemminger, "Acceptance Testing and Quality Control of Digital Mammography Equipment," in *Digital Mammography—Nijmegen, IWDM 1998 4th International Workshop on Digital Mammography*, edited by N. Karssemeijer, M. Thijssen, J. Hendricks, and L. van Erning (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998), pp. 437–444.
- ⁵M. J. Yaffe *et al.*, "Quality control for digital mammography: Part II recommendations from the ACRIN DMIST trial," *Med. Phys.* **33**, 737–752 (2006).
- ⁶M. Mahesh, "Digital mammography: An overview," *Radiographics* **24**, 1747–1760 (2004).
- ⁷I. Elbakri, A. Lakshminarayanan, and M. Tesic, "Automatic exposure control for a slot scanning full field digital mammography system," *Med. Phys.* **32**, 2763–2770 (2005).
- ⁸K. Fetterly and B. Schueler, "Performance evaluation of a dual-side read dedicated mammography computed radiography system," *Med. Phys.* **30**, 1843–1854 (2003).
- ⁹E. Berns, R. Hendrick, and G. Cutter, "Performance comparison of full-field digital mammography to screen-film mammography in clinical practice," *Med. Phys.* **29**, 830–834 (2002).
- ¹⁰C. Kimme Smith, C. Lewis, M. Beifuss, M. William, and L. Bassett, "Establishing minimum performance standards, calibration intervals and optimal exposure values for a whole breast digital mammography unit," *Med. Phys.* **25**, 2410–2416 (1998).
- ¹¹R. Saunders, Jr., E. Samei, J. Jesneck, and J. Lo, "Physical characterization of a prototype selenium-based full field digital mammography detector," *Med. Phys.* **32**, 588–599 (2005).
- ¹²S. Z. Shen, G. E. Mawdsley, and M. J. Yaffe, "Automation of Quality Assurance of Digital Mammography," in *IWDM 2000 5th International Workshop on Digital Mammography* (Medical Physics Publishing, Madison, 2000), pp. 656–665.
- ¹³M. J. Yaffe, D. B. Plewes, G. E. Mawdsley, J. P. Critten, S. A. Feig, A. D. A. Maidment, M. F. Picarro, and M. Tesic, "Full field digital mammography—System development and technical performance," in *Digital Mammography'96*, edited by K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt (Excerpta Medica International Congress Series 1119, 1996), pp. 125–132.
- ¹⁴J. P. Critten, K. A. Emde, G. E. Mawdsley, and M. J. Yaffe, "Digital mammography image correction and evaluation," in *Digital Mammography'96*, edited by K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt (Excerpta Medica International Congress Series 1119, 1996), pp. 455–458.
- ¹⁵C. B. Caldwell and M. J. Yaffe, "Development of an anthropomorphic breast phantom," *Med. Phys.* **17**, 273–280 (1990).
- ¹⁶J. F. Grey *et al.*, "Test pattern for video display and hard copy cameras," *Radiology* **154**, 519–527 (1985).
- ¹⁷J. C. Dainty and R. Shaw, *Image Science* (Academic Press, Toronto, 1974).
- ¹⁸P. S. Rezendes, A. de Almeida, and G. T. Barnes, "Mammography grid performance," *Radiology* **210**, 227–232 (1999).
- ¹⁹J. S. Bendat and A. G. Piersol, *Random Data Analysis and Measurement Procedures*, 2nd ed. (Wiley and Sons, New York, 1986), p. 285.
- ²⁰H. Fujita *et al.*, "A simple method for determining the modulation transfer function in digital radiography," *IEEE Trans. Med. Imaging* **11**, 34–39 (1992).
- ²¹D. R. Dance, C. L. Skinner, K. C. Young, J. R. Beckett, and C. J. Kotre, "Additional factors for the estimation of mean glandular breast dose using the UK mammography dosimetry protocol," *Phys. Med. Biol.* **45**, 3225–3240 (2000).
- ²²R. E. Hendrick, and E. A. Berns, *Full-Field Digital Mammography Quality Control Manual* (Phantom Image, Chicago, 2001).
- ²³K. D. Compton, "Factors affecting CRT display performance: Specifying what works," *Proc. SPIE* **3976**, 412–422 (2000).
- ²⁴National Electrical Manufacturer's Association, *Dicom Imaging and Communication in Medicine (DICOM) Part 14: Grayscale Standard Display Function*, PS 3.14–2004 (National Electrical Manufacturer's Association, Rosslyn, VA, 2004).
- ²⁵H. Blume, "ACR/NEMA Working Group XI: The ACR/NEMA proposal for a gray-scale display function standard," *Proc. SPIE Medical Imaging* **2707**, 344–360 (1996).
- ²⁶P. G. J. Barten, "Physical model for the contrast sensitivity of the human eye," in *Human Vision, Visual Processing and Digital Display III*, edited by B. Rogowitz (SPIE Press, Bellingham, WA, 1992), Vol. 1666, pp. 57–72.
- ²⁷P. G. J. Barten, "Spatio-temporal model for the contrast sensitivity of the human eye and its temporal aspects," in *Human Vision, Visual Processing and Digital Display IV*, edited by B. Rogowitz and J. Allebach (SPIE Press, Bellingham, WA, 1993), Vol. 1666, pp. 2–14.
- ²⁸See EPAPS Document No. E-MPHYA6-33-027602 for more detailed tables of results for many of the tests. This document can be reached via a direct link in the online article's HTML reference section or via the EPAPS homepage (<http://www.aip.org/pubservs/epaps.html>).