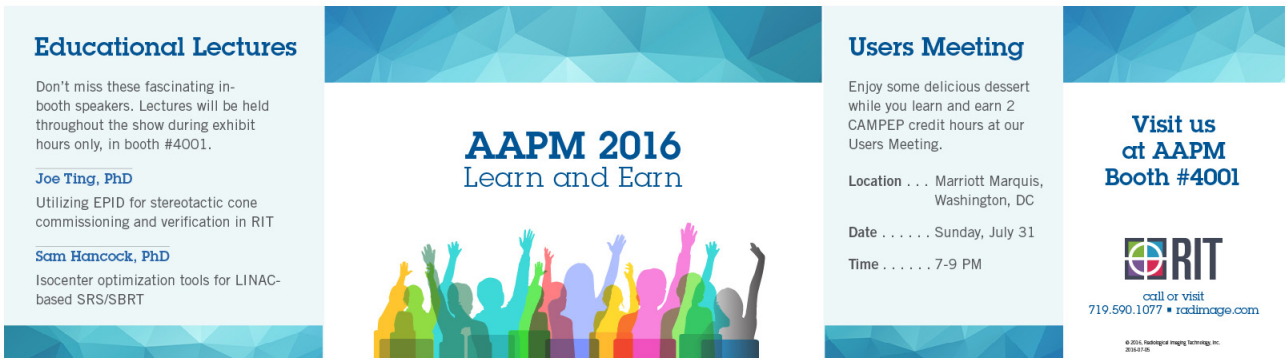
Medical Physics
The International Journal of Medical Physics Research and Practice

Quality control for digital mammography: Part II recommendations from the ACRIN DMIST trial

Martin J. Yaffe, Aili K. Bloomquist, Gordon E. Mawdsley, Etta D. Pisano, R. Edward Hendrick, Laurie L. Fajardo, John M. Boone, Kalpana Kanal, Mahadevappa Mahesh, Richard C. Fleischman, Joseph Och, Mark B. Williams, Daniel J. Beideck, and Andrew D. A. Maidment

Citation: Medical Physics **33**, 737 (2006); doi: 10.1118/1.2164067
View online: http://dx.doi.org/10.1118/1.2164067
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



Quality control for digital mammography: Part II recommendations from the ACRIN DMIST trial

Martin J. Yaffe, Aili K. Bloomquist, and Gordon E. Mawdsley
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

Laurie L. Fajardo
Department of Radiology, University of Iowa Health Care, 200 Hawkins Drive—3966 JPP, Iowa City, Iowa 52242-1077

John M. Boone
Department of Radiology and Biomedical Engineering, University of California Davis, Sacramento, California 95817

Kalpana Kanal
Department of Radiology, University of Washington, 1959 NE Pacific Street, Seattle, Washington 98195-7115

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

Richard C. Fleischman
Department of Medical Physics, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10021

Joseph Och
Allegheny General Hospital, Medical Physics, 320 East North Avenue, Pittsburgh, Pennsylvania 15212

Mark B. Williams
Radiology, Biomedical Engineering and Physics, University of Virginia, Box 801339, Charlottesville, Virginia, 22908

Daniel J. Beideck
Radiology Department, Fletcher Allen Health Care, University of Vermont, 111 Colchester Avenue, Burlington, Vermont 05401

Andrew D. A. Maidment
Department of Radiology, Hospital of the University of Pennsylvania, 1 Silverstein, 3400 Spruce Street, Philadelphia, Pennsylvania 19104

(Received 14 March 2005; revised 2 December 2005; accepted for publication 5 December 2005; published 23 February 2006)

The Digital Mammography Imaging Screening Trial (DMIST), conducted under the auspices of the American College of Radiology Imaging Network (ACRIN), is a clinical trial designed to compare the accuracy of digital versus screen-film mammography in a screening population [E. Pisano *et al.*, ACRIN 6652—Digital vs. Screen-Film Mammography, ACRIN (2001)]. Part I of this work described the Quality Control program developed to ensure consistency and optimal operation of the digital equipment. For many of the tests, there were no failures during the 24 months imaging was performed in DMIST. When systems failed, they generally did so suddenly rather than through gradual deterioration of performance. In this part, the utility and effectiveness of those tests are considered. This suggests that after verification of proper operation, routine extensive testing would be of minimal value. A recommended set of tests is presented including additional and improved tests, which we believe meet the intent and spirit of the Mammography Quality Standards Act regulations to ensure that full-field digital mammography systems are functioning correctly, and consistently producing mammograms of excellent image quality. © *2006 American Association of Physicists in Medicine*. [DOI: 10.1118/1.2164067]

Key words: digital mammography, quality control, image quality

I. INTRODUCTION

Digital mammography is an evolving imaging modality, quickly moving into regular clinical use. There are now several technologies available on the market, some of which are approved by the Food and Drug Administration for routine use in the U.S. At the present time, Mammography Quality Standards Act (MQSA) regulations require that sites follow the quality control (QC) procedures described by the individual manufacturers of the full-field digital mammography (FFDM) systems.[1] This has resulted in discordance among the QC protocols of the various systems.[2] To ensure that image quality is optimal and to support an effective accreditation program, routine QC, standard physics evaluation methods, and acceptance test practices that are independent of the manufacturer are required. With a view towards developing such routines and methods, this work took advantage of a unique opportunity to collect and analyze a significant amount of QC data from a large number of institutions and for all commercial FFDM units used in the American College of Radiology Imaging Network's (ACRIN) Digital Mammography Imaging Screening Trial (DMIST).[3]

In the DMIST trial it was considered imperative that the image quality of both screen-film mammography (SFM) and FFDM be representative of the full potential of each modality so that this would not be called into question after the trial. At the same time, there was very little experience available regarding the performance of FFDM systems. Therefore, the QC program for the trial, which we have described in Part I of this work,[4] was designed to be as comprehensive as possible based on a protocol developed for the International Digital Mammography Development Group (IDMDG)[5] and modified, so that as much as possible, tests could be applied generically among the different FFDM systems. In addition, however, in order to meet regulatory requirements it was necessary to incorporate all tests required by the specific equipment manufacturers. Because little was known regarding the expected modes or frequencies of equipment failure, a test schedule was designed with more frequent evaluations than that required for SFM systems.

In designing the QC tests for DMIST, we attempted to take advantage of opportunities for improvement of QC testing because of the availability of image data in digital form. This greatly facilitates computer analyses of images and allows for the introduction of objective and quantitative tests and more sophisticated measurements that are not practical for SFM (analog) systems.

Five different FFDM systems from four manufacturers were used in the DMIST trial. These are described in Part I, Table I. It should be noted that the DMIST facilities were considered to be demonstration sites for the equipment, and it should be expected that the companies would demonstrate extra vigilance to ensure that their units performed consistently. For some of the units, problems and design flaws were detected early in the program and rectified, and thus should not occur in future products. The Lorad Digital Breast Imager (Lorad DBI), with an array of CCD detectors, and the Fischer Senoscan I systems are no longer manufactured, so tests specific to those units are only discussed where relevant.

TABLE I. Failure rates for the congruence of the x-ray field and the field indicator.

Unit	Failures/Tests	Failure rate (%)
Fischer	12/61	20
Fuji	1/74	1
GE	1/130	1
Lorad DBI	0/9	0
Lorad Selenia	3/24	12
All systems	*17/298*	*6*

II. MOTIVATION AND RATIONALE

For a QC program to be practical and able to be followed by all facilities, some pragmatic decisions about the usefulness of individual tests and scope and extent of site survey testing must be made. It was found that the testing process was quite time consuming and that, while some of the information was relevant to the initial characterization of digital systems, it was of limited use for QC purposes. In addition, if one test can act as a surrogate for a number of others (offering high sensitivity, but possibly low selectivity), that test should be used first, and only if the system fails the initial test should it be necessary for the physicist or service person to perform further, more selective diagnostic tests.

In this work, the tests used in DMIST are considered in three categories depending on whether they evaluate the performance of the image acquisition system, the dose and image quality, or the image display system. The objective of each test is reviewed briefly and the pass/fail criteria used in DMIST are presented. Based on experience from the use of the test in DMIST, the utility of the test is discussed and modifications to the method of carrying out the test, including its elimination from the program and/or changes to the pass/fail criteria, are recommended. Tests which are to be performed annually are to be performed as part of the equipment evaluation or acceptance testing, which will provide reference (baseline) values for comparison on subsequent annual surveys. The intent is to develop a harmonized set of tests that could replace the different manufacturers' QC programs, and also allow for cross-vendor validation of system compatibility.

A. Failure rates

To assess whether tests were useful in identifying problems of imaging performance, we monitored the failure rates for each test. Certain tests had very low failure rates; this indicates that either there were few problems, or the limits were too lax, or the test was not discriminative of performance. These failure rates were based on the pass/fail criteria established for each test. The sensitivity of the failure rates to the pass/fail thresholds was also examined for some of the tests.

III. TESTS, DMIST RESULTS AND RECOMMENDATIONS

A. X-ray production and physical safety

In most radiological QC programs, emphasis is placed on the measurable parameters of the x-ray production system, and on basic dosimetry. At the time of development of the ACR Mammographic Accreditation Program for SFM, x-ray generator technology was rather simple and fluctuations in the quantity or quality of x rays produced were not uncommon. In particular, x-ray output was quite likely to drift over time. This could have an impact on image quality or radiation dose received by the breast. Modern x-ray generators used in digital systems, on the other hand, employ high-frequency technology and extensive feedback and control systems, ensuring that their performance is stable and well regulated. Furthermore, modern radiographic equipment performs internal self-tests and has interlocks that prevent exposures being initiated when problems are detected.

In SFM imaging, the optical density (OD) of the film and apparent contrast give indications that the system performance has changed. With digital imaging and the associated software manipulation performed by the unit, many problems can be masked, and the appearance of the image may not signal significant changes in the equipment operation.

1. Unit evaluation—breast thickness accuracy, maximum compression force, viewing conditions, etc.

Objective: To ensure that all locks, detents, angulation indicators, and mechanical support devices for the x-ray tube and breast support assembly are operating properly and that the DICOM header information is correct. The overall safety of the equipment is verified, and problems that might interfere with general operation are detected. A nonexclusive list of items to be checked regularly by the technologist covers most areas of outwardly observable physical faults. The physicist does a more thorough evaluation.

Pass/fail criteria: A number of the items on the list are subjective with suggested performance targets. Evaluation requires diligence and discernment on the part of the technologist. Where tests have objective measures, pass/fail criteria are similar to those for SFM systems through MQSA.

DMIST results: There were very few significant problems found at physics inspections, but it should be noted that these facilities were highly motivated, and had regular reinforcement of QC policies. The most common failure identified by physicists was the absence of posted technique charts. This was justified in most cases because the computer displayed a recommended technique, and there were no manual techniques used at those facilities. The most common problems seen by the technologists were those related to viewing conditions and monitors.

Utility: Regular checks by the technologist ensure that monitors are appropriately cleaned and that viewing conditions are appropriate. Some mechanical problems were found and service visits were scheduled to prevent downtime.

Recommendation: We recommend that this test set be retained, and that it be performed weekly by the technologist. The physicist should perform a thorough inspection on each annual visit. Additional items which were not included in the DMIST program include checking DICOM image header compliance for proper labeling, date, time, and time zone, all of which might get changed when software is upgraded. It is especially important that indicated breast thickness is accurate, as this affects technique selection and resultant breast dose.

2. CR Imaging plate fogging

Objective: To confirm that computed radiography (CR) plates are not fogged by radiation in their storage location.

Pass/fail criteria: There should be no evidence of fogging on the image. The shadow of a coin taped to the front of a cassette and left in place through a full day of imaging should not be visible even at the narrowest display window setting.

DMIST results: No evidence of imaging plate fogging was seen.

Utility: There is a low probability of failure.

Recommendation: This test is not recommended to be performed for routine QC.

3. Collimation and alignment

Proper collimation of the x-ray field is necessary to ensure there are no unexposed portions of the image receptor and that patients are not needlessly exposed to stray radiation. Proper alignment of the edge of the compression paddle with the chest-wall edge of the image-receptor holder assembly is necessary for proper positioning and compression of the breast.

Note that for CR, the image receptor is used with a number of different x-ray units and it is the alignment of the unit that is being evaluated in the following tests.

a. X-ray field, field indicator and image field congruency.

Objective: To evaluate whether the field as indicated by the machine (positioning light or other indicator) matches the true x-ray field and whether the x-ray field is congruent to the image receptor.

Pass/fail criteria: The sum of the x-ray field-indicator misalignments in the left-right and anterior-posterior directions should not exceed 2% of the source-to-image receptor distance (SID). The x-ray field should cover the entire displayed area, and no edge of the x-ray field should extend beyond the image receptor by more than 2% of the SID. Additionally, the x-ray field must not extend beyond the shielded area provided by the breast support except at the chest wall side.

DMIST results: Failure rates for the congruence of x-ray field and field indicator are shown in Table I, and that of the x-ray field and the image receptor shown in Table II. The high failure rates for the Fischer system arise from discrepancies between the points when x-ray exposure begins and ends during the scan (adjustable by the service engineer) and the field markings printed on the tabletop at the factory

TABLE II. Failure rates for the congruence of the x-ray field and the image receptor.

System	Failure/Tests	Failure rate (%)
Fischer	11/61	18
Fuji	5/74	7
GE	1/118	1
Lorad DBI	1/10	10
Lorad Selenia	2/24	8
All systems	*20/287*	*7*

(Table I) and the times at which image data collection begins and ends (Table II). There is no radiation hazard associated with this failure, provided that the x-ray detector or surrounding shielding absorbs the full area of the primary x-ray beam.

Utility: Accurate indication of the active image area is necessary for correct patient positioning. From experience with SFM systems, this test is useful for detecting gross errors in collimation adjustment or damage to the collimator device.

Recommendation: We recommend that the collimation tests be performed annually, and whenever major components that could affect alignment of collimation (x-ray tube, collimator parts, detector assembly, scanning drive) are repaired or replaced. The MQSA requirements for collimation can be met by all currently available systems evaluated in this study. In the future, this test will need to be done with a fluorescent screen, self-developing film, or an electronic "edge of field" imager because most facilities will not have access to a film processor.

b. Compression paddle and image receptor—Excluded tissue at chest wall.

Objectives: To ensure that the compression paddle is in the appropriate position and to determine the amount of tissue that is not imaged by the mammographic unit when a patient is positioned as closely to the unit as possible. A simple device[6,4] was imaged to assess the amount of tissue excluded from the image at the patient's chest wall.

Pass/fail criteria: The edge of the paddle is not to be visible in the image. Not more than 7 mm should be excluded.

DMIST results: Failure rates for the amount of excluded tissue are shown in Table III.

Utility: On some units, the paddle extension is adjustable, and improper alignment could result in poor positioning of the breast. If the edge of the compression paddle extends too far beyond the image receptor edge, the patient's chest is pushed away from the image receptor and some breast tissue will be excluded from the image. If the edge of the compression paddle does not extend far enough, the breast tissue will not be properly pulled away from the chest wall, resulting in poor compression at the chest wall, and the vertical edge of the compression paddle could obscure clinical information. Mechanical support structures or clearance for the chest-wall edge of the detector may result in unimaged tissue.

TABLE III. Failure rates for the excluded tissue test using more than 7 mm as the fail threshold.

System	Failures/Tests	Rate (%)
Fischer	11/35	31
Fuji	3/28	11
GE	0/57	0
Lorad DBI	6/8	75
Selenia	2/13	15
All systems	*22/141*	*16*

Recommendation: It is recommended that the collimation test be performed annually and following service to the x-ray tube or collimator or whenever the alignment of the breast support to the detector is adjusted. The 7 mm limit for missing tissue was found to be satisfactory in that all systems could be adjusted to achieve compliance.

4. kV Accuracy and reproducibility

Objective: This test evaluates the kilovoltage provided by the generator.

Pass/fail criteria: The measured kV must be within 5% of the nominal kV and the coefficient of variation (COV) between four successive exposures at the same kV setting must be less than 0.05.

DMIST results: Only one measurement exceeded the 5% limit.

Utility: Given the stability of modern x-ray generators, kV accuracy and reproducibility need not be tested as part of routine QC. Furthermore, noninvasive test instruments estimate kilovoltage based on beam quality and are less precise than voltage meters that are connected directly to the generator circuitry. Measurements of the half-value layer (HVL) of the x-ray beam will detect any gross problems with kV output, but for this to be effective as an alternative to measurement of kV, it is necessary to have more stringent criteria for the value of HVL and its consistency over time.

Recommendation: It is recommended that a measurement of HVL be used as an assessment of beam quality. kV should not be measured as a routine practice, but only by a service engineer using appropriately calibrated equipment at installation and when the generator is serviced.

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

Objective: This test ensures that tube x-ray output rate, linearity, and reproducibility meet MQSA requirements over a range of clinically relevant settings of kV, x-ray target, and beam filter.

Pass/fail criteria: For generator linearity, the output (mR/mAs) measured across a range of mAs settings was required in DMIST to remain within 10% of the mean tube output and increase monotonically with increased kV settings. The exposure output rate at 28 kV for those systems with Mo/Mo target filter combinations was required to be at least 800 mR/s (7 mGy/s air kerma) as specified by MQSA for

SFM systems. Output reproducibility requires the COV for four successive exposures to be less than 0.05.

DMIST results: There were five failures of tube output linearity in 143 testing instances. Of these failures, four occurred on Fischer systems and one occurred on a conventional unit being used with the Fuji CR system. Two of the four failures on the Fischer system were attributable to measuring the output at low mAs settings, well below the manufacturer's current recommended range of operation. One of the failures on the Fischer system, as well as the one on the Fuji system, is consistent with an operator data transcription error. None of the measurements of tube output (335 tests), output rate (102 tests), and output reproducibility (137 tests) indicated a failure (Sec. VA5 of Part I[4]).

Utility: Modern x-ray generators used in FFDM systems are universally of high-frequency design and incorporate internal feedback and correction circuitry that maintain virtually constant kV and mA during exposures. Exposure time is also controlled electronically and is highly reliable. X-ray tube output was found not to vary over long time periods.

Recommendation: We believe that it is still worthwhile to include the measurement of x-ray tube output under different tube target/filter/kV combinations as part of a routine QC program as an overall performance check, and also because these data are necessary for computing estimated mean glandular dose. However, all current mammographic x-ray sources easily meet the requirement for output rate, and performance of that test is not recommended. Testing mR/mAs and HVL will provide a warning of generator and spectral problems and will prompt diagnostic testing. It is not recommended to directly evaluate output linearity and exposure linearity. Image noise tests will provide a surrogate test for problems related to linearity and/or reproducibility.

6. Detector linearity and reproducibility

Objective: To evaluate the linearity of the detector response, the ratio of mean pixel value (MPV) to measured entrance exposure is tested for constancy.

Pass/fail criteria: The acceptance criterion for detector linearity is that at any point over a range of mAs (with other technical parameters constant), this value does not vary from its mean by more than 10%. For CR plates, sensitivity is deemed to pass if the S-number is within 15% of the target value. The limit on the COV for detector reproducibility measurements is 0.05.

DMIST results: The linearity and reproducibility of the detectors was found to be excellent for all systems. For the Fuji system, which has a logarithmic response, accurate S and L values are required to obtain linear results.

Of 136 tests of detector linearity, only seven failed—one on a Fischer system, two on Fuji systems, and four on Selenia systems (after offset correction had been applied). One of the failures on the Fuji system was due to a faulty photomultiplier tube in the CR reader.

There was only one instance out of 136 tests where the short-term reproducibility of the detector exceeded the COV limit of 0.05. The slight change in sensitivity did not significantly affect clinical image quality, and immediate service to the unit was not necessary.

Utility: Tests of linearity and reproducibility are helpful for characterizing detector response, but once the mammography unit is installed are of marginal utility.

Recommendation: We suggest that it should not be necessary to measure short-term detector reproducibility and linearity as part of routine QC. Instead, detector linearity and reproducibility should be measured only as a diagnostic tool, when irregularities are observed (i.e., shift of measured MPVs or "*S*-numbers") on signal measurements obtained from the technologist's weekly uniform phantom image. Logging of this information could be automated and unacceptable deviations could be used to trigger a warning message, prompting investigation of whether the deviation arose from the x-ray generation system or from the detector.

7. Half-value layer (HVL)

Objective: This test evaluates the effective energy of the x-ray beam. The HVL of the x-ray beam should be high enough to avoid excessive dose to the breast, while not so high that subject contrast is reduced to an unacceptable degree. The test also ensures that the x-ray beam quality is consistent with the target, filter, and kV selected, and enables the calculation of mean glandular dose.

Pass/fail criteria: At a given kV setting in the mammographic kilovoltage range (below 50 kVp), the measured HVL with the compression paddle in place must be within the range set out in the ACR Quality Control Manual.[7] For the upper limit, additional values of the constant, c, have been defined. For W/Mo target/filter combination, $c=0.28$ and for W/Al, $c=0.32$.

DMIST results: Of 396 measurements of HVL, no failures were recorded. The half-value layer showed little variation for any of the units.

Utility: If the HVL for SFM units is excessive, subject contrast will be reduced. For FFDM, with the capability of contrast manipulation, minor changes in HVL will have much less impact on image contrast than with SFM systems. Nevertheless, if kV is not measured routinely, HVL provides a check for variations in beam quality. In addition, knowledge of the HVL is required to estimate mean glandular dose.

Recommendation: This test should be performed annually. The HVL should be evaluated for each filter and for at least one kV that is typical of clinical operating techniques. We recommend that HVL tables should be provided by the manufacturer for each FFDM model to facilitate dose calculations and to allow verification of correct HVL. These tables should specify the expected HVL under typical target/filter/kV combinations for clinical use. After initial testing, if the HVL is compliant with the manufacturer's specification, the measured value should be adopted as the reference value and changes from that value tracked. The currently permitted range of HVL is very wide, and designed to accommodate a range of equipment designs; sensitivity to changes in HVL will be easier to detect if there is an established operating

point. In the DMIST measurements, it was found that HVL varied typically by no more than 3%. A requirement that HVL be constant within 6% seems to be reasonable.

8. Focal spot

Objective: This test ensures that the spatial distribution of x-ray emission from the focal spot in contact or magnification mode does not unduly degrade spatial resolution of the image.

Pass/fail criteria: The limiting effective spatial resolution in line pairs/mm for a bar pattern, placed 4.5 cm above the breast support table, was measured on a mammographic SFM receptor placed on the breast support surface. The MQSA SFM criteria were employed; the minimum required limiting resolution was 11 line pairs/mm with the pattern bars perpendicular to the anode-cathode axis and 13 line pairs/mm with the pattern bars parallel to the anode-cathode axis. Where magnification capability was available, systems were tested according to the same criteria using the small focal spot with the resolution pattern placed 4.5 cm above the magnification stand.

DMIST results: All but 5 of 246 measurements (2%) met the MQSA requirements. Of those that failed, three were taken using the magnification stand, and were just outside the limits. There was no measurable reduction of *overall* spatial resolution on mammographic units which failed this test.

Utility: For contact mammography, the effective resolution provided by the focal spot is considerably higher than the resolution limit imposed by the digital detector and therefore has little influence on the overall MTF of the system. A test of system MTF is considered to be a more sensitive, objective, and relevant measurement of resolution, and is capable of detecting problems with the focal spot, synchronization errors, and other factors affecting system resolution. As more facilities switch to digital imaging, the film and processors needed to make this test will become unavailable.

Recommendation: The separate measurement of focal spot resolution using SFM images need not be performed, except as a diagnostic test to evaluate resolution problems detected using the MTF test.

B. Image quality and radiation dose

The measurement of image quality in mammography has been a longstanding challenge. Many objective test techniques for assessing the physical variables of imaging such as spatial resolution, contrast, noise, and dynamic range are available. In most cases, however, the link between these measures and clinical image quality has not been solidly established. Nevertheless, there is good reason to accept that there is a relationship between the above-mentioned variables and both subjective image quality as assessed by a radiologist and diagnostic accuracy (sensitivity and specificity).[8–13] The assessment of image quality in FFDM is further complicated by the image processing applied in FFDM. This can in some instances compensate for limitations or abnormalities in the image acquisition stage; however, it can also introduce artifacts or possibly suppress important image information. The testing procedures employed in DMIST attempted to combine objective physical measures with subjective assessment of phantom images.

1. Daily accreditation phantom imaging

Objective: This test ensures that equipment operating characteristics have not changed, and that there are no obvious artifacts in the images.

Pass/fail criteria: Images of the mammography accreditation program phantom (MAP) were scored centrally by a trained reader using the ACR guidelines,[7] and the minimum passing score was set to match the MQSA requirement of 4 fibers, 3 speck groups and 3 masses.

DMIST results: Over 5766 images were scored by a single reference reader. Only 20 images represented the types of image quality problems that would require subjective assessment of the visibility of test objects in a phantom. The other 37 failures were clearly caused by technical problems (incorrect technique selection, blank images, and severe artifacts) that would be visible in a uniform phantom image.

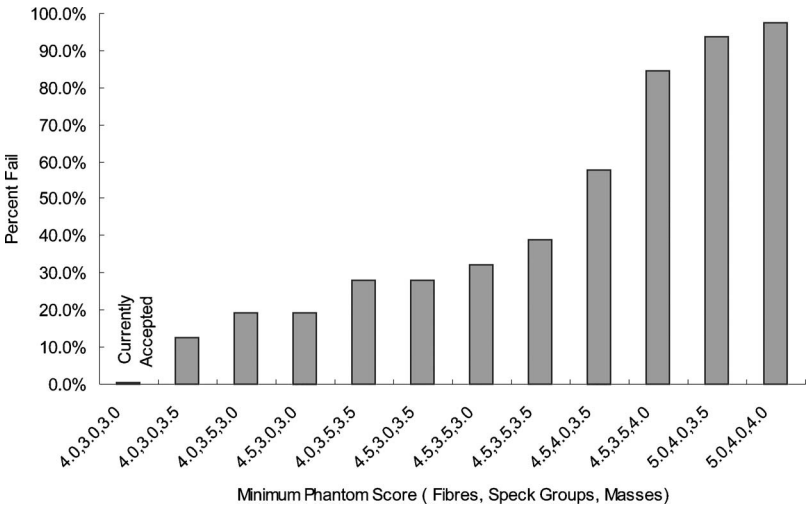


FIG. 1. Illustrative failure rates calculated by applying more stringent mammography accreditation phantom threshold scores.

TABLE IV. Failure rates for technologists scores of weekly accreditation phantom images.

System	Failures/*N*	Failure rate (%)
Fischer	3/1091	0.3
Fuji	0/59	0.0
GE	0/1116	0.0
Lorad DBI	0/118	0.0
Lorad Selenia	1/243	0.4
All systems	*4/2627*	0.15

Utility: The current SFM failure limits allowed all digital units to pass, suggesting that this phantom has very little discriminative capability with FFDM. There is evidence that the phantom as manufactured has significant variability, and that scores for the same system can be different using different phantoms, and when performed by different readers.[14]

One option to improve the utility of this familiar test object could be to increase the thresholds for passing this phantom to reflect the adjustable contrast capability of FFDM. For example, the minimum passing threshold score could be raised from the current standard of 4 fibers, 3 speck groups, and 3 masses to other higher levels. The failure rates for different minimum threshold phantom scores are given in Fig. 1. The failure rate increases rapidly if any single threshold is adjusted, suggesting that the intervals between structures in the accreditation phantom are too coarse to allow detection of subtle problems or changes in image quality arising in FFDM.[15]

These observations motivate a shift to test objects amenable to automated quantitative analysis for FFDM.

Recommendation: The current mammography accreditation phantom, designed for SFM systems, is not discriminative enough to be appropriate for QC of FFDM systems. It would be valuable to develop a phantom that is more discriminative of image quality in FFDM, while still being capable of being scored in a reasonable amount of time and being as user independent as possible.

2. Weekly accreditation phantom imaging

Objective: This test is intended to ensure that the images being produced by the FFDM system are of acceptable quality. The image should be viewed on the soft-copy workstation or printed film, whichever method is used by the radiologist to read clinical images.

Pass/fail criteria: Weekly MAP images scored by the technologist should at least meet the SFM requirements.

DMIST results: The failure rates of the technologist scored MAP images for each system type are given in Table IV. On average, the technologists scored the phantom images about $\frac{1}{2}$ point higher for each test object than the reference reader.

Utility: None of the failures could be correlated with failures recorded for the phantom images scored centrally (only one image that failed was found in both databases, and its score from the central physicist passed).

Recommendation: Routine use of the current mammography accreditation phantom used for SFM systems should be eliminated for FFDM QC. More discriminative tests such as signal-difference-to-noise-ratio (SDNR, described below) artifact evaluation, and measurement of MTF should be used for routine QC, and clear guidance as to pass/fail criteria should be provided for the technologist.

3. Weekly imaging of uniform phantom

Objectives: This test is intended to demonstrate consistency in tube output, AEC operation, and detector operation as well as to detect image artifacts. Measurements of signal level (MPVs on soft-copy images or OD on printed films) and noise [standard deviation (SD)] are taken at given locations in the phantom image and compared against established baseline values. The mAs used to acquire the image is also tracked.

Pass/fail criteria: Variations in the signal level or OD, noise, and mAs of more than 10% from the established baselines were defined as "failures."

DMIST results: For this test, 1846 measurements by the site technologists were evaluated. The overall failure rate for

TABLE V. Failure rates for the weekly uniform phantom test for each system. The Fischer and Lorad Selenia systems used a fixed manual technique to image the phantom so the mAs used could not vary. The Fuji, Lorad DBI, and Lorad Selenia systems performed this test using a printed image, so noise was not evaluated. MPV =mean pixel value.

System	Phantoms imaged (*N*)	mAs failures (%)	MPV or OD failures (%)	Noise failures (%)	Overall failure rate (%)
Fischer	546	NA	0.37	4.4	4.6
Fuji	480	6.7	5.2	NA	11
GE	736	3.0	6.7	0.1	8.7
Lorad DBI	66	1.5	20	NA	21
Lorad Selenia	18	NA	11	NA	11
All systems	*1846*	*4.3*	*4.9*	*2.0*	*8.6*

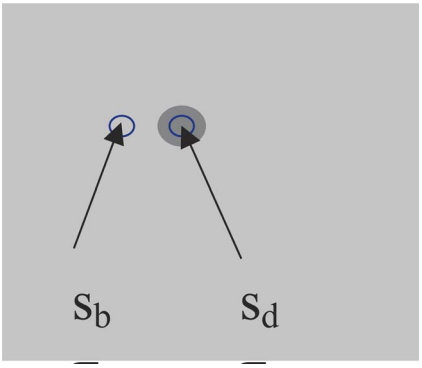


FIG. 2. Illustration of measurement of signal-difference-to-noise ratio (SDNR).

the weekly uniform phantom test was 8.6%. Failure rates for the different systems are given in Table V.

Utility: The failure rates of the weekly uniform phantom suggest the need for a test to track the performance of the system and ensure consistency. Imaging of a uniform phantom including one area with known and easily measured contrast will provide better consistency and reduced operator variability than can be obtained using the mammography accreditation phantom.

Recommendation: A phantom should be imaged weekly and examined for artifacts. Signal and noise measurements should be performed to verify consistent behavior of the imaging chain. A suggested metric is the signal-difference-to-noise ratio (SDNR). A simple test object consisting of a 4.0 cm slab of uniform attenuating material (PMMA) with a flat-bottomed 1 mm deep depression on its upper surface is imaged. Regions of interest of equal areas are selected in the image of the disk (d) and in an adjacent background region (b) (see Fig. 2). In each region the MPV, s, and the per-pixel SD about s, σ, are determined. The SDNR is given by

$$\mathrm{SDNR} = \frac{(s_d - s_b)}{(\sigma_d^2 + \sigma_b^2)^{1/2}}.$$

This test is intended for QC purposes and should provide a useful tool for monitoring changes in image quality. The measurement is affected, however, by the correlation of noise between pixels. Therefore, results will depend on the system MTF, so that SDNR alone is not a suitable tool for comparing different system types. For that purpose, measurement of the spatial frequency dependent noise-equivalent quanta [NEQ(f)] would be more appropriate.

4. Artifacts

Objective: To assess the degree and source of artifacts visualized in the digital image and to ensure that the flat-field image is uniform. For CR systems, the test evaluates the uniformity of the imaging plates, reader, and the printer/processor subsystem.

Pass/fail criteria: Under appropriate viewing conditions, using reasonable window and level settings (similar to those for clinical viewing, but with a slightly narrower window), there should be no visible dead pixels, missing lines, or missing columns of data. There should be no visually distracting structured noise patterns in an image of a uniform phantom. There should be no regions of discernibly different signal level (apart from heel effect) or OD on a displayed processed image.

DMIST results: A summary of the artifacts found is given in Table VI.

Utility: Subjective appraisal of the image of a uniform slab of plastic was found to be a most effective means of finding imaging system problems. Images should be viewed on hard copy or soft copy, as normally used for clinical work, with defined display settings that provide somewhat higher (but not excessively higher) contrast than is normally used for clinical viewing.

The visibility of artifacts depends on the contrast setting of the display system. If the contrast setting employed while inspecting for artifacts is unrealistically high compared to the settings used for clinical viewing, it is likely that artifacts will be noticed that would not normally impair lesion detection or characterization tasks. More work on determining an appropriate and reproducible method for displaying images to evaluate artifacts is required. For CR systems, where flat fielding is not performed, and therefore image nonuniformities due to phenomena such as the heel effect will be present, more subjective criteria similar to those used for SFM systems may be more appropriate.

Recommendation: The physicist should test for artifacts

TABLE VI. Number of artifacts found by physicists. N is the number of tests performed. Misc. indicates miscellaneous other artifact causes. Some images had multiple artifacts with multiple causes.

		Artifact cause								
System	N	Flat-fielding	Motion	Misc.	Filter	Ghosting	Bright/Dark pixels	Grid	CR reader	Images with artifacts (%)
Fischer	37	17	2	2	0	0	0	0	—	57
Fuji	29	9	0	6	1	0	0	4	4	72
GE	57	9	0	4	7	2	1	0	—	39
Lorad DBI	8	4	0	0	0	0	7	0	—	88
Lorad Selenia	13	2	0	0	0	0	6	0	—	62
All systems	*144*	*41*	*2*	*12*	*8*	*2*	*14*	*4*	*4*	*55*

annually. This is complementary to the weekly test done in conjunction with the SDNR measurement by the technologist. The use of a uniform phantom image for the detection of artifacts is probably the most effective test for the maintenance of high-quality imaging. Since an effective flat-fielding algorithm can hide many problems with individual detector elements or even rows of data, information on the location of "bad" pixels and image rows, or a "dead pixel map" should be obtained. Thresholds for acceptable numbers of bad pixels need to be determined as a percentage of the total image size, but more importantly, the nature of these imperfections (clustered, adjacent rows or columns, etc.) must be specified so that the significance of these artifacts on clinical image quality can be assessed.

5. Misty/conspicuity test

Objective: To evaluate the ability of the system to demonstrate low contrast objects and fine detail.

Pass/fail criteria: Because of lack of *a priori* experience with the imaging systems used in DMIST, there were no pre-established limits for this test.

DMIST results: While this test was qualitatively interesting, the readings were highly operator dependent, time consuming, and subjective. There was no consistent pattern seen between scores and MTF, kV, or entrance exposure.

Utility: This test is of marginal utility as a QC test.

Recommendation: As used in DMIST, the Misty phantom was not sufficiently discriminative of differences in image quality and overall system performance. Therefore, we do not recommend using it for routine FFDM QC. While a test of the ability of the overall system to render subtle anatomical details visible is desirable, we are not confident that any existing phantom can be evaluated in a reliable manner that will distinguish optimal from suboptimal performance in FFDM. Further work is necessary either in phantom design or in definition of methods for consistent evaluation of images.

6. Noise levels and noise power spectrum

a. Noise vs signal level.

Objective: To evaluate the spatial and electronic noise characteristics of the entire imaging chain.

Pass/Fail Criteria: For nominally linear systems, regions of interest (ROI) of 4.0 cm^2 distributed over the area of the image were required to have an R^2 (coefficient of determination) value greater than 0.95 for a linear, least-square fit to variance (SD squared) vs signal level (MPV). For CR (nominally logarithmic) systems, the same minimum R^2 value was required for a linear least-square fit to variance vs the "S-Number" (inversely proportional with exposure). If one area is not linear and displays an excess of noise, this region will limit the allowable operating range of the system. Any significant change should be evaluated and corrected.

DMIST results: All systems showed a strong linear relationship between variance and exposure, indicating that the noise is close to being quantum limited.

Utility: Systems with malfunctions causing increased electronic noise (e.g., a defective photomultiplier tube), showed reductions in R^2. This test is useful to characterize the performance of the image acquisition subsystem.

Recommendation: This test should be performed annually, and after servicing of the detector or digitization subsystems.

b. Nonrandom noise.

Objective: To determine the amount of nonrandom (structured) noise present in images.

Pass/fail criteria: The SD of the pixel values in a ROI placed within an image computed as the average of four images acquired with nominally identical exposure factors should be approximately half of the corresponding SD calculated from one of those images. Larger values indicate the presence of significant structured (nonrandom) noise, prompting investigation.

DMIST results: The Fuji system, which does not incorporate uniformity correction for the x-ray field or plates, demonstrated the highest amount of nonrandom noise.

Utility: This test was found to be a good objective method of evaluating images for structural or nonrandom variations.

Recommendation: This test should be performed upon acceptance testing of the unit, and after servicing of the detector or digitization subsystems.

c. Noise power spectrum (NPS)

Objective: To characterize the spatial frequency content of image noise.

Pass/fail criteria: There were no established criteria or limits set for DMIST, although if a significant spike was detected, the system underwent further analysis.

DMIST results: The most frequently observed phenomenon was the presence of discrete spikes in the NPS. These occurred at spatial frequencies corresponding to the interline spacing of the grid when one was used. However, there were virtually no periodic structures observed in those images.

Utility: It was challenging to establish a universal metric for evaluating and comparing the power spectra because of uncertainty in the spectra themselves (proper measurement of noise power requires many replicate measurements) and difficulties in normalizing signal levels between systems. Blotches or small single-point artifacts do not have enough power to demonstrate a measurable change in the NPS. Because of these factors, routine measurement of noise power spectrum was not useful.

Recommendation: The performance of NPS is not recommended as a routine test; however, it may be helpful in diagnosing problems identified by the SDNR test, or when spatially repetitive artifacts are observed such as those caused by improper grid motion or textures in structural components of the system (e.g., breast support surface).

7. Effective system modulation transfer function

Objective: To determine the modulation transfer function (MTF) for the overall imaging system.

Pass/fail criteria: Because this was a new test for FFDM

systems and because we had limited experience with the performance of different systems, we did not set pass/fail criteria at the onset of the study.

DMIST results: Typical MTF results for each system type are described in Part I.[4] Since no pass/fail criteria existed during the trial, failure rates for this test are not provided.

Utility: With images available in digital format and user-friendly software, it is straightforward to perform this test. This is in contrast to the effort and precision required to measure MTF on SFM systems. This test ensures that hardware is performing properly and is not degrading the resolution of the image below original equipment performance levels. It provides an estimate of the effective detector element (del) aperture size, rather than the nominal value based on the spacing between image samples. This test is extremely important for systems with moving parts or with read-out systems where the aperture and sampling pitch can vary.

Recommendation: It is recommended that MTF be measured annually, and after service to the detector, tube, bucky, or CR plate reader. The MTF of the system in the magnification configuration should also be measured. For systems with moving parts in the image chain (i.e., scanning systems or CR), it is recommended that MTF also be tested monthly by the technologist. To facilitate use by the technologist, software for the calculation of MTF should be developed that incorporates the pass criteria and communicates the pass/fail result clearly to the user.

Mechanical motions in scanning systems used for either image acquisition or readout can also affect the MTF. Scanning systems require that the speed of the scanner be maintained at a constant value, so for mechanically scanning acquisition systems, MTF should be checked at gantry positions of −90, 0, and +90 deg.

While an absolute requirement on MTF might be appropriate at some future time, we do not currently know what MTF is required to adequately detect and diagnose breast cancer. Image quality factors such as noise and image processing applied subsequent to acquisition interact with MTF in defining clinical image quality. Therefore, at this time we recommend that the required MTF be specified relative to the expected performance as specified by the manufacturer. The minimum acceptable value of the MTF could be specified at different fractions of the Nyquist limit of the system (determined from the manufacturer's quoted del size). For example, the Nyquist limit of a system with dels at a 50 micron pitch is 10 cycles/mm and a minimum acceptable transfer ratio of 40% at 0.5 of Nyquist requires the MTF at 5 cycles/mm to be at least 40%.

In addition, the MTF should be isotropic, requiring that the MTFs calculated along the principal axes of the image differ by not more than 0.08 at a spatial frequency of 2 mm^{-1}, ensuring consistent quality in representing fine details regardless of orientation. This was twice the SD among the systems surveyed in the DMIST trial.

Since many of the systems perform image processing, increasing the visibility of some details, and possibly suppressing noise, methods to analyze the postprocessed image will need to be developed.

TABLE VII. Failure rates for thickness tracking test, by machine type.

System	Failures/Tests	Failure rate (%)
Fischer	0/35	0
Fuji	6/24	25
GE	0/52	0
Lorad DBI	0/6	0
Lorad Selenia	4/13	31
All systems	***10/130***	***8***

8. Thickness tracking

Objective: To evaluate the ability of the system to image a range of x-ray attenuations that simulates clinical breast imaging and to ensure that images of adequate penetration and acceptable signal and signal-to-noise ratio (SNR) levels are produced.

Pass/fail criteria: The passing criteria for this test were taken from the manufacturer's QC protocol, and varied with each system. The Fischer system required that the SNR (the ratio of the MPV in a ROI to the SD of the pixel values in the same region) be greater than 50 for all thicknesses. The GE system required that the SNR be greater than 50 for 2 and 4 cm thicknesses of specified attenuator and greater than 40 for 6 and 8 cm. The Lorad DBI system required that the ratio of the SNR for each thickness to the average SNR be between 0.80 and 1.20. The Selenia system required that the SNR be greater than 40. For the CR system, the S-numbers for the different exposures were required to be within 15% of their mean value.

DMIST results: The failure rates for the thickness tracking tests are given in Table VII. The lack of conformance seen in the test results for Fuji may be due to incorrect calibration of the mammography unit, use of phantoms that are too small, or incorrect positioning of the phantoms such that they are not located over the area being used by the CR processing algorithm to determine the S value. Three of the four failures on the Selenia units were due to faulty manual technique charts.

Utility: FFDM equipment can provide viewable images over a wide range of breast doses. Image processing operations can be used to smooth noise and amplify contrast, and in so doing, may mask the use of an inappropriate radiographic technique. Without tracking, imaging performance and dose might change, with hardly perceptible changes in clinical images. Optimal performance of any of the FFDM systems over the range of breast thicknesses and compositions requires the use of an effective automatic technique selection method.

This test allows tracking of the relationship between the average image pixel value and the radiation exposure to the breast for different degrees of x-ray attenuation, simulating changes in thickness and/or composition of the breast. How-

ever, simply ensuring that the MPV remains approximately constant may not ensure that adequate SNR is maintained. Failure of the technique selection controls of the system to respond to changes in thickness or composition could result in ineffective use of the dynamic range of the digital system and cause a decrease in the SNR. Currently, there is no consistent policy between manufacturers for setting the AEC target level.[4]

Recommendation: The performance of a thickness tracking test is useful; however, we believe it can be more useful if it incorporates an additional measure related to subject contrast, i.e., the signal difference. The SNR may not constitute a sensitive enough indicator of system performance, since it does not include any assessment of true radiographic signal, which is an indication in the image of differences in x-ray attenuation between rays through the breast. The SDNR provides an index of image quality that relates to the ability to discern subtle structures in the breast in the presence of noise. The metric $\mathrm{SDNR}/(\mathrm{dose})^{1/2}$ provides a measure of the dose efficiency of the imaging system.[16]

More experience is necessary to allow determination of the recommended range of SDNR for different thicknesses. This test should be performed at least annually.

9. Geometric distortion

Objective: To determine the absolute image magnification and the fidelity with which straight lines are imaged.

Pass/fail criteria: No visible local blurring, bending of the lines, or discontinuities in the structures in the test tool shown in Fig. 4 in Part I was permissible.

DMIST results: During DMIST, machines that did not utilize mechanical scanning (i.e., those other than the Fuji and Fischer systems) did not demonstrate changes in distortion between inspections. The Lorad DBI demonstrated gaps between the fiber-optic tapers, and localized pin-cushion distortion. The Fischer system showed synchronization-related artifacts and minor gaps between the fiber-optic tapers.

Utility: This test is capable of detecting areas of insensitivity between sections of the detector as well as speed synchronization problems in systems that employ mechanical scanning (including CR). This test is also useful for checking the accuracy of annotation tools and determining the magnification factor of the displayed image compared to the actual breast size.

Recommendation: It is recommended that this test be done only on acceptance and after service to the detector assembly, except for systems with mechanical scanning of the x-ray source and/or detector and on CR systems, which employ a laser scanner, where annual testing by the physicist is advisable.

10. Entrance exposure and mean glandular dose

Objective: To measure the typical entrance exposure for a "standard" breast (approximately 4.2 cm compressed breast thickness—50% adipose, 50% fibro-glandular composition), and to calculate the associated mean glandular dose (MGD).

Pass/fail criteria: The mean glandular dose to a standard breast should not exceed 3 mGy (0.3 rads) per view. Note that in some jurisdictions, the limit is 2 mGy. If the values exceed these levels, action should be taken to evaluate and eliminate the cause of excessive dose.

DMIST results: The estimated average mean glandular doses for the standard breast were either equivalent to or below those given for film-screen systems, and well within the 3.0 mGy limit. The CR exposures varied more than the dedicated FFDM systems because existing SFM units were used for CR, and the AEC systems on those units were programed to match the preferences of the individual sites' radiologists using the local screen-film combination.

Utility: It is a fundamental principle of radiation safety that the user be aware of the amount of radiation used to produce a mammogram. Underexposure can result in inadequate SDNR which may cause a reduction in image quality, while overexposure unduly exposes the patient and may saturate the detector.

Recommendation: This test should be performed annually by the medical physicist. We recommend that if the MGD is displayed for an image, or reported in the DICOM header, that value should match the value calculated by the physicist to within 15%.

11. Image detector ghosting/lag

Objective: To evaluate the severity of any artifact due to previous exposure to the detector. In this measurement, a ghost or residual image induced in a similar manner to clinical use is quantified.

While a test for ghosting was not initially included in the DMIST test protocol, it was discovered that a number of systems demonstrated subtle after-images of previously imaged objects on the succeeding uniform flat-field images. After investigation, it was found that a number of the digital detectors demonstrated "ghosting," a loss of detector sensitivity in areas that had previously received high exposure. In the worst case, the ghost artifact represented a 10% loss of sensitivity on one of the Lorad Selenia units. This sensitivity loss was the result of cumulative exposure to the unattenuated x-ray beam in regions outside the breast during routine patient imaging.

A new test was developed to evaluate this phenomenon in terms of the reduction of the MPV produced by the detector per incident exposure.

DMIST results: Under typical exposure conditions for the standard breast, the maximum sensitivity loss observed due to a single exposure was 1.3%. The artifact was not considered to be of clinical significance but could be noticed if the image was viewed with very narrow window settings, particularly near the periphery of the breast.

Pass/fail criteria: It will be necessary to determine pass/fail criteria when there is more experience with this test.

Recommendations: A reasonable test for this artifact has been incorporated into the European Reference Organisation for Quality Assured Breast Screening and Diagnostic Services' (EUREF) FFDM QC program.[17] The test involves measuring the change in detector sensitivity due to ghosting

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

	Fischer	GE	Selenia
N (number of tests)	31	60	7
Failure rate (%)	0	3	0

under controlled conditions, and comparing the change in sensitivity to a known signal contrast. We recommend that the test described in the EUREF QC program be performed on all types of systems annually and upon replacement of the detector.

C. Display

The interface of the FFDM acquisition system to the diagnostician through the physician review station is probably the most important link in the FFDM chain. Analogous to SFM, where the processor and viewboxes can be extremely variable, the digital display devices, both soft copy and hard copy, are susceptible to misadjustment and drift and are often ignored. Picture archiving and communications systems (PACS) are becoming common, and the ability to display images acquired by multiple devices produced by different vendors is very desirable. If there are multiple locations where primary diagnosis is performed, all of those devices must meet standards. It is important to test each device (monitor or film printer) with images in exactly the format that is provided in the output of the acquisition device, because there are variations in implementation of the DICOM Standards.

1. Monitor evaluation

A digital test pattern was displayed on the soft-copy workstation. Subjective tests of spatial resolution, contrast, and artifacts were carried out and quantitative measurements of brightness in test areas in the pattern were made.

a. Overall display quality—SMPTE pattern.

Objective: To ensure that the displayed image is a true representation of the "for presentation image" as represented by a standard digital test image in terms of contrast rendition over a range of brightness, spatial resolution, and freedom from artifacts.

Pass/fail criteria: All gray-level steps of the Society of Motion Picture and Television Engineers (SMPTE) pattern should be distinct from one-another; All line-pair patterns over the image field should be resolved; The 0%–5% and 95%–100% contrast patterns should be visible. No excessive blurring, streaking, smearing, or other artifacts should be present. The overall pass or fail score for each monitor was left to the physicist's discretion.

DMIST results: The overall failure rates for review workstation monitors when evaluating the SMPTE pattern are given in Table VIII. These results differ from those of Table XII in Part I because the physicist chose to pass several monitors despite failure of individual SMPTE test components, having decided that these were of minor concern, or a consequence of system design.

Utility: The SMPTE pattern provides a good sense of monitor performance. However, we found that the AAPM Task-Group 18 QC test pattern,[18] which became available after our study was underway, provides a more thorough means of verifying correct monitor performance because it provides a more comprehensive set of contrast steps over the full range of monitor driving levels. The full set of TG-18 tests is very labor intensive, and not appropriate for field surveys.

Recommendations: Because the monitor is now the primary means for interpretation of digital mammograms for many systems, some degree of monitor testing should be performed at least weekly by the technologist. A more thorough characterization of monitor performance, including resolution testing, should be performed by the physicist at least annually, given that the luminance output of cathode ray tube (CRT) monitors, when operated at the high brightness conditions of mammography viewing, drops as the phosphor ages.[19]

We recommend that the TG18-QC test pattern should be displayed weekly and examined on all primary medical display devices used to interpret digital mammograms, using test images emulating (i.e., having the same *x-y* dimensions, number of bits, and a DICOM header containing appropriate values of all relevant tags) the images produced by each model of FFDM unit in the facility, or which might be interpreted at that workstation. The system should be evaluated under typical operating conditions and room lighting levels. It is essential that cross-vendor compatibility be verified before images from one vendor can be interpreted on another vendor's workstation. This includes any specialized software considered to be important for proper viewing of such images.

b. Monitor luminance response measurement.

Objectives: To evaluate how closely the gray-scale calibration of the monitor meets the DICOM gray-scale display function (GSDF) and to ensure that digital soft-copy review workstation monitors are of adequate brightness and contrast.[20] Conformance with the DICOM GSDF should ensure that the luminance response is perceptually linear and that images are displayed consistently.

Pass/fail criteria: Because of limited experience with soft-copy display, and the fact that not all manufacturers claimed that their workstations were calibrated according to the DICOM GSDF, no pass-fail criteria were set for this test.

DMIST results: The photometers used for this test had insufficient precision at low luminance levels, only measuring to the nearest candela/meter^2 (nit). Also, use of the small gray-level patches in the SMPTE pattern to measure luminance response may have resulted in glare from surrounding areas affecting the measurement accuracy. This problem could have been circumvented by using the "zoom" tool to enlarge the patches for measurement, but this was not done in our tests. Good conformance to the DICOM GSDF was possible with both GE and Fischer review workstations. Lorad did not claim that its workstation was calibrated to be perceptually linear, and in our tests it was found to provide a

markedly different luminance response, as detailed in Part I. Fuji did not provide a soft-copy display option for DMIST.

Recommendation: Monitors should be visually inspected for correct calibration by checking the contrast of the TG-18 QC test pattern at least weekly. The luminance response of monitors should be measured by a physicist at least annually using the TG18 protocol. For this purpose, the TG18-LN patterns provide a series of 18 squares covering the full range of driving levels. Each square is centered within a larger background square at mid-gray level. The autocalibration software and self-monitoring features that are now often supplied with newer monitors should make maintaining correct monitor calibration less onerous; however, it is important to verify luminance with an independent photometer in case the one attached to the unit becomes inaccurate.

The photometer used to evaluate luminance response should be accurate to within 5% over a range of 0.05 to 1000 nits and provide a precision of at least 1% of reading.[18]

We note that the complete TG18 test program for monitors is extremely extensive and provides excellent tools for the laboratory environment. Flat panel and LCD monitors will require a different testing protocol than that used for CRT monitors, and we expect that TG-18 will evolve to meet these requirements. For practical clinical field testing, we believe that an abbreviated version of that program is acceptable. This should include qualitative evaluation of the TG18 QC pattern and a reduced set of spatial resolution measurements.

2. Laser printer evaluation

a. Printer calibration and artifacts.

Objectives: To evaluate how closely the OD of the film meets the DICOM gray-scale display function (GSDF) and that the printer produces consistent images of high quality.

Pass/fail criteria: For the SMPTE pattern, similar criteria were applied as those for the monitor. All density steps should be distinct and the 5% squares at both bright and dark levels must be visible when the film is placed on the radiologist's mammographic viewbox with appropriate masking. No pass-fail criteria were set for the conformance of the printers to the DICOM GSDF. There were to be no significant artifacts noted in the printed film of a uniform image.

DMIST results: There were no failures of the printed SMPTE pattern seen in 41 test instances. The best-calibrated systems conformed quite well to the DICOM GSDF. Overall conformance to the standard was not impressive, but conformance was not explicitly required at the beginning of the study. Of 67 printer evaluations, there were only two occasions when the physicist failed the printer. Both printer failures were due to excessive artifacts seen on images printed on the Agfa LR5200, which prompted servicing of the equipment.

Utility: Adherence to the DICOM GSDF standard will help to ensure that printed images have a similar appearance to those viewed on soft-copy display systems, and provide consistent results across multiple vendor systems. This test is a good method of determining the gray-scale display function of the printer.

TABLE IX. Sensitometry failure rates for the different printers. *N* is the number of data points. B+F is base plus fog. LD is low density. MD is mid-density. DD is density difference.

Printer	*N*	B+F or LD (%)	MD (%)	DD (%)
Agfa 4500	104	2.9	51.0	39.4
Agfa LR5200 (wet process)	1315	2.1	3.4	2.4
Fuji FM-DPL	2332	3.0	2.0	2.3
Kodak 8610	2018	0.0	1.1	1.4

Recommendations: This test should be performed annually on all printers used to print digital mammograms. Printing should be done from the review workstation, so that any image transformations performed during the image transfer and printing processes are included. The TG18-QC pattern[18] suggested for monitor evaluation in Sec. IIIC1a is recommended for this purpose. Further experience is needed to determine a reasonable criterion for determining acceptable conformance to the DICOM GSDF. The technologist should visually inspect the printed test pattern quarterly to ensure that high printed image quality is maintained. In addition, a uniform image should be printed monthly to verify that artifacts are minimal.

b. Printed MAP image.

Objective: This test provides a subjective evaluation of printed image quality.

Pass/fail criteria: The printed film image of the MAP was required to demonstrate at least 4 fibers, 3 specks, and 3 masses.

DMIST results: None of the printed MAP images evaluated by the physicists had failing phantom scores.

Utility: This test provides no additional information about printer performance.

Recommendation: This test should not be performed as part of routine FFDM QC. The appropriate TG-18 QC pattern mimicking the "For Presentation" image format of the acquisition device should be printed and evaluated instead.

c. Printer sensitometry.

Objective: This test monitors the stability and reproducibility of the laser printer.

Pass/fail criteria: Criteria are similar to the film sensitometry procedure followed under MQSA. The OD of the mid-density step and density difference should be within 0.15 OD units of their target values. The measured base plus fog should be no more than 0.03 OD units above the target value.

DMIST results: The failure rates for the different aspects of the laser printer sensitometry are given in Table IX. The printers were quite stable and reproducible, especially the dry printers. The high failure rates for the one site with the AGFA 4500 may be due to incorrect target values being set by the local technologist, as the measured mid-density and density-difference values were quite stable, as shown in Part I.[4]

Utility: Wet processing of films is subject to many detri-

TABLE X. Recommended FFDM technologist's tests—QC procedures and minimum frequencies.

Test name	FFDM system	Minimum frequency
Viewing conditions	All	Daily
Monitor cleaning	All	Daily
Laser printer sensitometry check	All hard copy	Daily or weekly
Darkroom/printer area cleanliness	All hard copy	Daily
Phantom image quality (SDNR and artifact evaluation)[a]	All	Weekly
Review workstation monitor	All	Weekly
Acquisition monitor	All	Weekly
Viewbox cleanliness	All hard copy	Weekly
MTF	All with moving parts in image chain	Monthly
Mechanical inspection	All	Monthly
Laser printer artifacts	All hard copy	Monthly
Repeat analysis	All	Quarterly
Laser printer—TG18 pattern[a]	All hardcopy	Quarterly
Compression force	All	Semiannually

[a]Test revised from initial DMIST protocol.

mental external influences such as variations in solution concentrations, temperature, and emulsion formulation. For dry processing, there are fewer variables, and the systems have the capability of performing self-sensitometry.

Recommendation: For printers with wet processing, the sensitometry test should be performed daily. For printers with dry processing, stability is greater and the test could be performed weekly. It may be possible to further reduce the frequency in the future, as more experience is obtained and printer technology matures.

3. Soft-copy viewing conditions assessment

Objective: To assure that the ambient diffuse light levels (illuminance) incident on the review work station monitors do not degrade the quality of the clinical images.

Pass/fail criteria: Light incident on the monitor surface is not to exceed 10 lux.

DMIST results: None of the facilities exceeded the limit.

Recommendations: The ambient room illuminance falling on the monitor must be measured by a medical physicist at least annually or when lighting changes are made. Because the light incident on the monitor degrades the perceived contrast in displayed images, illuminance levels on the monitor should be no greater than the level used in this study and should be maintained at the same level as when the monitor was calibrated to the DICOM GDSF. In addition, no specular reflections should appear on the monitor screen. The ambient light level recommendations and protocols in AAPM TG-18

TABLE XI. Recommended FFDM physicist's tests—QC procedures and minimum frequencies.

Test name	FFDM system	Minimum frequency
Mammographic unit assembly evaluation	All	Yearly
Artifact evaluation	All	Yearly
Ghost image evaluation[a]	All	Yearly, replacement of detector
Breast entrance exposure and mean glandular dose	All	Yearly
Chest wall missed tissue	All	Yearly, servicing, detector replacement
Modulation transfer function (MTF) measurement	All	Yearly
X-ray field evaluation	All	Yearly, servicing
Tube output—mR/mAs vs kV	All	Yearly
Beam quality assessment—HVL	All	Yearly
Technique chart/AEC evaluation (SDNR)[b]	All	Yearly
Noise and linearity	All	Acceptance, servicing
Spatial linearity and geometric distortion of the detector	All	Yearly, servicing
Monitor display quality[b]	All soft copy	Yearly
Monitor resolution[b]	All soft copy	Yearly
Monitor luminance response and viewing conditions[b]	All soft copy	Yearly
Viewbox luminance and room illuminance	All hard copy	Yearly
Laser printer evaluation[b]	All hard copy	Yearly

[a]Additional test, see Sec. III B 10.
[b]Test revised from initial DMIST protocol.

report should be followed. A daily room lighting check list should be available to the technologists which provides them guidance in reviewing the lighting conditions to assure that they are maintained in an approved manner.

4. Viewbox luminance and illuminance of the reading room

Objectives: To ensure that the luminance of the viewboxes used for both interpretation and quality checking of printed mammograms meets or exceeds minimum levels, room illuminance levels for both soft and hard copy reading are sufficiently low, and viewing conditions have been optimized.

Pass/fail criteria: ACR requirements for SFM were applied.

DMIST results: There were no failures reported. The sites used their own light meters to measure viewing room illuminance. Viewbox measurements were done for MQSA reports. All sites were in compliance, as determined from the local physicists' reports.

Recommendations: As with SFM interpretation, proper viewing conditions are important to ensure that visualization of details is possible across the full dynamic range of the images. To comply with the DICOM GSDF, the calibration of the film printer must be done with knowledge of the luminance of the viewboxes on which the resulting films will be interpreted and the ambient illuminance, as these affect the perceived luminance levels in the resulting films. Thus, once a printer is correctly calibrated, it becomes important to ensure that viewing conditions do not change significantly. As is the case with workstations, the minimization of extraneous light is very important. The technologist should verify that lighting conditions in the reading room are acceptable daily. The physicist should measure viewbox luminance and reading room illuminance annually.

IV. RECOMMENDATIONS FOR QC TESTING PROTOCOL

The DMIST study provided an opportunity to explore which QC tests would be valuable for FFDM. Based on our DMIST experience, we can make some recommendations regarding test equipment, the types of tests that are useful, and their frequencies.

A. Suggested test objects

The phantoms used in DMIST were designed to be reproducible in small batches, and to allow us to determine which test objects were most useful. While subjective tests are appealing, they tend to be poor indicators of more objective measures of system performance, since it is very difficult to replicate the critical tasks in breast imaging and difficult to evaluate phantom images in a consistent manner both within and between observers. For this reason, we recommend the use of quantitative tests that are objectively evaluated, except for the assessment of artifacts.

Our three recommendations regarding test devices are

(1) A uniform flat phantom with a 1 mm deep flat-bottomed well and reference target objects should be used to verify that artifacts are minimal and permit a measurement of the signal-difference-to-noise ratio. We believe that this will be the best practical indicator of image quality and equipment performance. The reference objects are structures that can be viewed to aid in setting display levels for evaluating image artifacts, and assessing artifact severity.
(2) For measurement of MTF, a 25 mm medium-contrast square with sharp edges should be used[21] to record the edge-spread function, from which the MTF can be calculated. The pattern should be positioned at the level of the upper surface of the standard breast. The test would be facilitated by the provision of validated software with a user-friendly interface for both the physicist and the technologist.
(3) The TG18 QC and TG18 LN patterns with appropriate format for the individual FFDM acquisition systems should be used for evaluation of soft and hard copy displays.

B. Summary of recommended tests and suggested frequencies

From the extensive testing performed in DMIST and our analysis of test results, we have been able to formulate a recommended QC program for FFDM as described in previous sections. This is summarized in Table X and Table XI. Many of the test methods are described in Part I, and as much as possible, these tests are similar to those familiar to technologists and medical physicists who carry out QC in SFM.

V. CONCLUSIONS

Review of the physics QC data from the DMIST program suggests that some of the tests in the program are useful, either in their original form or with slight refinements, for evaluating system performance or predicting the probability of failure, while others provided little useful information. In addition, certain currently performed tests, primarily tests on the x-ray generator function, appear to be of very little value in FFDM, suggesting that they could be removed from a periodic testing regimen.

Even though DMIST was a large imaging trial, in some cases the number of times a particular QC test was performed and the time span of the testing was not large enough to establish statistically meaningful thresholds of acceptable performance or mean times between failures. For practical purposes, there must always be compromises between the time required to perform tests, and the degree of characterization of the system that is achieved. Digital imaging systems lend themselves to quantitative, automated testing procedures which are self-logging. We recommend that these testing procedures be implemented to as great an extent as practical. This will contribute to high compliance in QC test-

ing while reducing the impact on both cost and the time of valuable personnel. Efforts are now underway to standardize algorithms for calculation of imaging metrics such as MTF and SDNR. When these algorithms are available, we recommend that the manufacturers of FFDM systems include them in a QC module that is part of the system to facilitate consistency in testing.

In some cases, the test procedures have been modified based on the DMIST experience. There is still a need to further improve and streamline test procedures for the display monitors.

We believe that these recommendations will provide a useful framework for definition of a QC program for FFDM. Without question, recommended QC procedures will evolve along with the systems, and with our increasing understanding of FFDM. It will be necessary to modify the tests, their frequency, and pass/fail criteria as more experience is gained in the field, and as the technology matures. We are optimistic that a more generalized and less labor-intensive harmonized QC program can be developed based on this knowledge.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the American College of Radiology Imaging Network, funded by NIH/National Cancer Institute, and the collaboration of the local physicists and technologists at all sites involved in the Digital Mammography Imaging Screening Trial. Development of the QC tests was in part supported by the Ontario Research and Development Challenge Fund.

[1]MSQA. Mammography Quality Standards Act Regulations 21 CFR PART 900—Mammography (2002).

[2]M. B. Williams, P. J. Goodale, and P. F. Butler, "The current status of full-field digital mammography quality control," JACR **1**, 936–951 (2004).

[3]E. Pisano, R. E. Hendrick, S. Masood, and C. Gatsonis, *American College of Radiology Imaging Network—ACRIN 6652—Digital vs. Screen-Film Mammography*, ACRIN (2001).

[4]A. Bloomquist *et al.*, "Quality control for digital mammography in the ACRIN DMIST trial: Part I," Med. Phys. **33**, 719–736 (2006).

[5]G. 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.

[6]J. 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.

[7]R. 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).

[8]A. G. Haus and M. J. Yaffe, "Screen-film and digital mammography—Image quality and radiation dose considerations," Bur. Stand. (U.S.), Bull. **38**, 871–898 (2000).

[9]A. G. Haus, M. J. Yaffe, S. A. Feig, R. E. Hendrick, P. A. Butler, P. A. Wilcox, and S. Bansai, "Relationship between phantom failure rates and radiation dose in mammography accreditation," Med. Phys. **28**, 2297–2301 (2001).

[10]K. C. Young, M. G. Wallis, and M. L. Ramsdale, "Mammographic film density and detection of small breast cancers," Clin. Radiol. **49**, 461–465 (1994).

[11]L. W. Bassett, D. M. Farria, S. Bansal, M. A. Farquhar, P. A. Wilcox-Buchalla, and S. A. Feig, "Reasons for failure of a mammography unit at clinical image review in the American College of Radiology Mammography Accreditation Program," Radiology **215**, 698–702 (2000).

[12]C. J. Vyborny and R. A. Schmidt, "Technical image quality and the visibility of mammographic detail," in *A Categorical Course in Physics: Technical Aspects of Breast Imaging 3rd edition*, edited by A. G. Haus and M. J. Yaffe (RSNA Publications, Oak Brook, IL, 1994), pp. 103–110.

[13]L. N. Loo, K. Doi, and C. E. Metz, "A comparison of physical image quality indices and observer performance in the radiographic detection of nylon beads," Phys. Med. Biol. **29**, 837–856 (1984).

[14]K. Imamura *et al.*, "Synchrotron radiation imaging showed cracking—like structures in ACR-approved mammography phantoms," Med. Phys. **30**, 1457 (abstract TH-C24A-8) (2003).

[15]D. P. Chakraborty, "Computer analysis of mammography phantom images (CAMPI): An application to the measurement of microcalcification image quality of directly acquired digital images," Med. Phys. **24**, 1269–1277 (1997).

[16]M. B. Williams *et al.*, "Beam Optimization for Digital Mammography," in *IWDM 2000 5th International Workshop on Digital Mammography* (Medical Physics Publishing, Madison, 2000), pp. 108–119.

[17]European Commission, *The European Protocol for the Quality Control of the Physical and Technical Aspects of Mammography Screening, Part 2B Digital Mammography*, 4th ed. (in press, 2006).

[18]E. Samei *et al.*, "Assessment of Display Performance for Medical Imaging Systems," Med. Phys. **32**, 1205–1225 (2005).

[19]K. D. Compton, "Factors affecting CRT display performance: specifying what works," Proc. SPIE **3976**, 412–422 (2000).

[20]National Electrical Manufacturer's Association, *Digital Imaging and Communication in Medicine (DICOM) Part 14: Grayscale Standard Display Function*, PS 3.14-2004 (National Electrical Manufacturer's Association, Rosslyn, VA, 2004).

[21]A. K. Carton *et al.*, "Validation of MTF measurement for digital mammography quality control," Med. Phys. **32**, 1684–1695 (2005).