

Premarket Applications for Digital Mammography Systems; Final Guidance for Industry and FDA

Document issued on: February 16, 2001

**This document supersedes Information for Manufacturers Seeking
Marketing Clearance of Digital Mammography Systems: Status Update,
February 4, 1999**



**U.S. Department Of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health**

**Radiology Devices Branch
Division of Reproductive, Abdominal, and Radiological Devices
Office of Device Evaluation**

Preface

Public Comment

Comments and suggestions may be submitted at any time for Agency consideration to Dockets Management Branch, Division of Management Systems and Policy, Office of Human Resources and Management Services, Food and Drug Administration, 5630 Fishers Lane, Room 1061, (HFA-305), Rockville, MD, 20852. When submitting comments, please refer to the exact title of this guidance document. Comments may not be acted upon by the Agency until the document is next revised or updated.

For questions regarding the use or interpretation of this guidance contact John C. Monahan at (301) 594-1212 or email jcm@cdrh.fda.gov.

Additional Copies

Additional copies are available from the Internet at: <http://www.fda.gov/cdrh/ode/guidance/983.pdf> or CDRH Facts-On-Demand. In order to receive this document via your fax machine, call the CDRH Facts-On-Demand system at 800-899-0381 or 301-827-0111 from a touch-tone telephone. Press 1 to enter the system. At the second voice prompt, press 1 to order a document. Enter the document number 983 followed by the pound sign (#). Follow the remaining voice prompts to complete your request.

Table of Contents

Purpose	4
Introduction	4
The Least Burdensome Approach	5
Background	5
Regulatory Requirements	6
Non-clinical Information	6
A. Device Characteristics.....	7
B. Performance Standards.....	8
C. Technical Data.....	8
Clinical Information	11
A. Trial Designs.....	12
B. Presentation of Clinical Data.....	14
C. Post-approval Unenriched Screening Study.....	14
Hard- and Soft-copy Display	16
Labeling	16
A. Indications for Use.....	16
B. Contraindications.....	16
C. Warnings/Precautions.....	16
D. Adverse Events.....	17
E. Summary of Non-clinical Studies.....	17
F. Summary of Clinical Studies.....	17
G. Instructions for Use.....	17
Quality Assurance Program	18
MQSA Requirements	18
Appendices	20
I. Definition of Terms.....	20
II. Approval of Soft-copy (Hard-copy) Display Subsequent to Hard-copy (Soft-copy).....	22
III. ROC Methodology.....	24

Premarket Applications for Digital Mammography Systems; Final Guidance for Industry and FDA

This document is intended to provide guidance. It represents the Agency's current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind the Food and Drug Administration (FDA) or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statute and regulations.

Purpose

This document is intended to provide guidance to industry on the type of information needed by the Center for Devices and Radiological Health (CDRH) to evaluate a marketing application for a full-field digital mammographic (FFDM) system.

Introduction

The primary use of mammography is in the screening and diagnosis of breast cancer. Digital mammography systems are intended as a replacement for analog (film-screen) systems for all mammographic uses, i.e., screening as well as diagnostic including problem solving (See **Appendix I** for definitions of terms used in this document.). As such, market entry for digital mammography systems requires more extensive testing than does market entry for digital imaging systems for purely diagnostic radiographic devices, e.g., for chest, bone, etc.

The clinical impact of any screening modality lies in the severity of the condition being screened for, the impact of early detection, and the number of people affected. Unlike a diagnostic modality, which affects only those who are suspected of having the particular disease in question, screening is applied to all those persons who are deemed to be at risk of developing the disease, the vast majority of whom, however, are not suspected of having it. For a low prevalence disease such as breast cancer, in which only about half a percent of those screened annually will be found to have the disease, the number of those not expected to have it is very large. Furthermore, because of the number of women at risk, the number of those who do actually develop the disease, though a few orders of magnitude smaller, is also quite large. In the case of breast cancer, the number of new cases each year is estimated to be approximately 180,000. Therefore, small increases or decreases in the performance of screening modalities will affect large numbers of individuals.

Given the minimal radiation risk of modern mammography, the main concerns regarding digital mammography relate to its ability to permit detection of abnormalities of the breast and accurate discrimination between malignant and benign findings. In particular, the parameters of interest are sensitivity (or its complement, false negative rate, i.e., how many cancers are delayed in their detection and diagnosis?) and specificity (or its complement, false positive rate, i.e., how many women are subjected to surgical procedures who turn out not to have cancer?).

A significant amount can be learned about the capabilities of digital mammography systems through assessment of the technical or physical characteristics of those systems; however, a clear correlation between design/bench parameters and clinical performance has yet to be established. The key element of clinical performance is the radiologist's report. The linkages between the bench performance of the acquisition component and the radiologist's report consist of display (hard- or soft-copy) and perception and interpretation by the radiologist. The combined effect of all these aspects is best evaluated through appropriate clinical trials.

The Least Burdensome Approach

The issues identified in this guidance document represent those that we believe need to be addressed before your device can be approved/cleared for marketing. In developing the guidance, we carefully considered the relevant statutory criteria for Agency decision-making. We also considered the burden that may be incurred in your attempt to comply with the guidance and address the issues we have identified. We believe that we have considered the least burdensome approach to resolving the issues presented in the guidance document. If, however, you believe that information is being requested that is not relevant to the regulatory decision for your pending application or that there is a less burdensome way to address the issues, you should follow the procedures outlined in the "A Suggested Approach to Resolving Least Burdensome Issues" document. It is available on our Center web page at: <http://www.fda.gov/cdrh/modact/leastburdensome.html>

Background

In 1996, based on consultation with industry, NIH, and the clinical community, including input from a meeting of the Radiological Devices Advisory Panel the year before, the FDA issued a guidance on the preparation of marketing submissions for digital mammography systems. The guidance suggested that an "agreement approach" to study design might result in the development of data that could be used to support a substantial equivalence determination and market entry through the 510(k) premarket notification pathway. The "agreement approach" represented an attempt to minimize the clinical data requirements that would have been necessitated by a controlled clinical study in a screening population requiring the enrollment of tens of thousands of women. A study designed to show agreement would involve a small fraction of that number and would have, therefore, been less burdensome.

However, subsequent attempts by manufacturers to demonstrate agreement between each woman's digital and analog mammograms (as outlined in the guidance) suggested that this approach was impractical, primarily due to the magnitude of inter- and intra-observer variability inherent in mammographic interpretation, and secondarily due to the intra-patient variability associated with repositioning necessary when the same breast was imaged on two separate mammographic devices. The data from the first 510(k) submitted demonstrated that the inter- and intra-reader variability was so great that agreement as defined in the 1996 guidance was not possible, for this or any other digital mammography device. The agency determined that the data was inadequate to determine equivalence and therefore a "not-substantially-equivalent" decision was made. As a result, this and all other devices within this category were automatically classified in Class III.

The agency has continued discussions regarding digital mammography with manufacturers, the medical community, academics, other federal agency staff, and other interested parties. FDA recognizes both the concern regarding trial sizes so large as to be prohibitive for individual manufacturers, and, at the

same time, the need for adequate data to provide the reasonable assurance of safety and effectiveness for each system, as mandated by Congress. The health of millions of U.S. women over 40 years of age, who depend on mammography for early detection of breast cancer, must not be adversely affected if digital mammography systems are to replace current analog systems. Taking these considerations into account, the agency is presenting a new approach to market entry for digital mammography systems, which, we believe, is feasible and still provides appropriate and sufficient information to understand their basic clinical performance characteristics.

Given the importance of mammography to public health, the complexities inherent in performing a valid comparison of analog and digital mammography, and our not-substantially-equivalent decisions, the PMA route to market was required. This process allows each manufacturer to establish safety and effectiveness for their own device. We believe that this is the “least burdensome” method for providing the valid scientific evidence needed to bring digital mammography to market. Our recent experience with the first PMA submitted under this paradigm which was filed, reviewed, taken to panel and approved within 91 days seems to have validated this approach. While we cannot guarantee that all subsequent applications will achieve the same successful outcome in the same time frame, we have both publicly and privately stated our commitment to work with individual manufacturers to assess their current situation, formulate a plan to maximally leverage the data which they already have, and require additionally only that information which is needed to provide a complete and potentially approvable PMA application.

Regulatory Requirements

Prior to marketing, a new mammographic technology must conform to regulations developed under three different laws and amendments to those laws: the Radiation Control for Health and Safety Act of 1968 (RCHSA); the Medical Device Amendments to the Food, Drug, and Cosmetic Act of 1976 (MDA); and the Mammography Quality Standards Act of 1992 (MQSA).

Since the x-ray part of a digital system may be unchanged from analog (e.g., the tube, generator, support, grid, beam-limiting device, etc.) and only the image receptor system is altered, conformance with the regulations of RCHSA represents little that is new and no special guidance is provided here.

Under the Medical Device Amendments, a device may be cleared for marketing via a 510(k) premarket notification or it can receive approval via the premarket approval application (PMA) route. When a device cannot be shown to be substantially equivalent to a legally marketed device (510(k)), the sponsor must establish the safety and effectiveness of the device for its intended use, (PMA).

Under MQSA, a wide range of quality assurance procedures for analog mammography devices have been developed to assure that the analog systems are operating at or above some base level of performance. The same types of tests need to be developed for digital mammography systems to assure that the system is operating properly and that it remains above some accepted baseline level of performance. Until such systems are in place, use of digital systems will be permissible only in facilities accredited for analog mammography.

Non-clinical Information

This part of the guidance describes the non-clinical content of submissions for x-ray equipment and accessories used for digital mammography. It should be noted that a digital mammography system

encompasses all aspects of the imaging process from data acquisition to image display. As such, the guidance addresses not only the x-ray system and data acquisition mechanism but also the image viewing device(s), whether hard- or soft-copy, along with any image processing used to prepare the acquired data for display. Manufacturers who choose to seek marketing approval for soft-copy (hard-copy) image display subsequent to approval of hard-copy (soft-copy) image display for their digital mammography system should refer to [Appendix II](#).

A. Device Characteristics

The device characteristics section of the premarket submission provides the basic information needed to give the reviewer a thorough appreciation of the technical specifications of the device and its components. The following information should be included:

1. A complete description of the software development cycle, which should include:
 - a. a description of algorithms used,
 - b. a hazard analysis,
 - c. the listing of software requirements,
 - d. a description of software with structure chart,
 - e. a summary of software development process,
 - f. a summary of verification and validation activities, and
 - g. a summary of verification and validation results.
2. A complete description of the entire system, including pictorial representations of the layout and interconnection of the different components.
3. A complete description of each of the functional components of the device including the x-ray system, the image acquisition and recording device, and the image viewing and display device(s).
4. A complete description of the properties of the device relevant to its physical capabilities. This description should include any appropriate technical characteristics and specifications not only for the entire imaging system but also for the components of the system, including:
 - a. x-ray tube: target(s) material, x-ray filter(s) type and thickness, window material and thickness, focal spot size(s)
 - b. x-ray generator: type, range and accuracy of technique factors (x-ray tube voltage, x-ray tube current, exposure time and mAs) for each focal spot size, if applicable
 - c. geometry: source to image receptor distance (SID), source to patient support device distance, alignment to chest wall
 - d. x-ray scatter grid: grid ratio, Bucky factor, stationary or reciprocating
 - e. x-ray detector: material, interaction efficiency, geometrical characteristics, optical path, scanning rate (for systems using a slot scanning system), and decay rate of the phosphor afterglow (for systems using a storage phosphor)
 - f. analog to digital conversion (ADC): bit depth, matrix size, pixel width
 - g. digital to analog conversion (DAC)
 - h. soft-copy display system: type, screen area (active video area, raster boundary), resolution (MTF, contrast modulation, spot size, video bandwidth), brightness (veiling glare, dynamic range), characteristic curve, number of addressable pixels (matrix size) and physical

- construction (dispenser cathode, deflection angle, phosphor type, antireflective coating)
- i. laser film recorder: type, processor, characteristic curve or sensitometric response (look-up tables), measure of resolution (MTF, contrast modulation, spot size), matrix size, format size, artifact rejection
 - j. image-processing algorithms: description, suitability, selection
 - k. image processing platform: type, word length, language
 - l. digital archiving system and data security considerations: access by unauthorized personnel, retention of case and patient information

Note: Technical product literature or brochures can be used to provide some or all of the above information.

5. A complete description of the device's principles of operation including:
 - a. discussion of the methods used to select the technique factors on the x-ray system
 - b. rationale for using any automatic exposure control systems for controlling the x-ray exposure
 - c. method of assessing and choosing among the available image processing algorithms
 - d. process for generating the display data from the detected data, and for viewing the display data, including:
 - for soft-copy display, general recommendations for an appropriate “hanging protocol” such as image orientation and location, initial viewing conditions (e.g., display function, views, ROI, etc.), and options for image display manipulation (e.g., window level, window width)
 - for soft- or hard-copy display, general recommendations for ambient light levels and viewing conditions when using the display device.

B. Performance Standards

The performance standards section of a premarket submission normally contains information concerning performance standards that are applicable to the device. In this regard, relevant information includes

1. certification that the device meets the requirements of the Radiation Control for Health and Safety Act of 1968

The name of any other applicable standard should be provided with a complete explanation for any deviation from the standard. Also, a clear and concise reference should be provided for any other submission to the FDA, such as product reports and supplemental reports, and abbreviated reports.
2. declaration of conformity to any other voluntary standards or data to demonstrate safety associated with electrical, mechanical, radiation, digital interface, software and display characteristics, e.g., “Digital Imaging and Communications in Medicine (DICOM) Part 14: Grayscale Standard Display Function”

C. Technical Data

The technical sections of a premarket submission normally provide the information needed to evaluate the validity and accuracy of non-clinical laboratory studies. The natural separation of detection and

display in the process of image forming and viewing in a digital imaging system warrants the splitting of non-clinical laboratory studies into those two categories. Relevant information includes the following

1. Quality of Detected Data:

a. quantum limited operation—

For quantum limited performance (i.e., noise added by the FFDM system does not exceed the quantum noise when operated in the normal range of exposures), provide data showing that the device operates in a quantum limited mode at the exposure levels specified for its use. If not quantum limited, provide the range of exposures where quantum limited operation is not achieved.

b. sensitometric response--quantitative measure of the sensitometric response of the image acquisition system (i.e., the digital value versus radiation exposure curve)

c. spatial resolution--quantitative measure of the spatial resolution properties of the image acquisition system (i.e., the modulation transfer function (MTF))

d. SNR transfer--quantitative measure of the efficiency of SNR transfer of the image acquisition system as measured by the noise equivalent quanta (NEQ) and/or the detective quantum efficiency (DQE) as a function of spatial frequency

For systems using flat-field correction, the impact of flat-field correction on DQE and NEQ.

e. dynamic range--quantitative measure of the dynamic range of the image acquisition system as measured by the NEQ and/or the DQE as a function of spatial frequency and radiation exposure level.

f. phantom images--description of test results using the phantom approved or accepted by FDA and any other phantoms deemed appropriate, such as a contrast-detail (CD) phantom.

g. image erasure/fading.

For systems using a delayed readout of image data such as a photostimulable phosphor, description of test results on image fading as a function of time and temperature, retention of image information as a function of the number of erasures and/or exposures, and information on fogging and depletion of charge after exposure to room light including results of fading test at 50°C, if the system is recommended for batch processing in a mobile facility.

h. a demonstration, on the basis of at least 100 repeated exposures and erasings, that there are no residual trapped charges that can give false information as multiple exposures or ghost images

Manufacturers should also indicate the life of a cassette and the criteria for replacement, i.e., after how many exposures a cassette must be replaced.

i. defect characteristics—

Describe the allowable types and quantities of defects, and any methods of compensation or blanking that are utilized to compensate for them. The description should include the methods used to correct for pixel-to-pixel variations in sensitivity, offset, etc.

- j. noise analysis—
Provide a quantitative measure of the noise properties of the image acquisition system (i.e., the noise power spectrum (NPS)) as a function of spatial frequency and exposure level. (Note: These data should form the basis for any statements associated with DQE, quantum limited operation and dynamic range requested elsewhere in this guidance.)
2. Quality of Displayed Data—for soft-copy display systems or hard-copy recorder, as appropriate
- a. Soft-copy display system.
Provide quantitative measures of the maximum luminance, the minimum luminance under recommended ambient light and viewing conditions, the luminance dynamic range, the gray scale display function, and the luminance uniformity and a description of the method of assuring adequate performance of the soft-copy device with respect to resolution, veiling glare and noise, including the basis for any specified level of performance and the type of test used to verify the level of performance.
 - b. Hard-copy recorder.
Provide quantitative measures of the sensitometric response (e.g., gamma, digital value versus brightness), a measure of spatial resolution (e.g., modulation transfer function, limiting resolution), a measure of artifactual and electronic noise (e.g., variance, NEQ, DQE, SNR transfer as a function of spatial frequency), and a measure of dynamic range.

Note: All of the data provided in parts 1 and 2 of section C should be supplemented with a complete description of all phantoms, test protocols, and digital mammography system settings used to determine the stated imaging performance. If the test object is readily available in the imaging community (e.g., the SMPTE test pattern), a simple reference would be sufficient. The supplemental information should also include a statement of the uncertainty in the stated imaging performance data.

3. Patient Radiation Dose

Provide a quantitative estimate of the patient radiation dose expressed as the average glandular dose delivered during a single cranio-caudal view of an accepted phantom, e.g., simulating 2, 4.2 and 6 cm thick, compressed breasts consisting of 70, 50 and 30 percent glandular and 30, 50 and 70 percent adipose tissue, respectively (9 measurements). The manufacturer should also provide a complete description of the conditions of operation on the mammography system, including but not limited to kVp, mAs, x-ray filtration, and exposure level at the detector. The patient radiation dose should be determined with the technique factors and conditions that are used to produce the images of the phantom approved or accepted by the FDA and any other phantoms deemed appropriate, such as a contrast-detail (CD) phantom.

4. Test Results

Provide representative images associated with the measurements described in parts 1 and 2 of this section. These images may be on film recorded from the output and/or other recording media including digital data in DICOM format on CD-ROM. Information on imaging system set-up including technique factors, geometry, focal spot, etc. and on any image processing applied to the digital data should be provided along with any representative images. The characteristics of any test patterns on imaging phantoms should also be provided, including but not limited to physical parameters such as the spatial frequency (lp/mm) of each test object for resolution gauges. In addition, provide representative data associated with the measurements described in parts 1 and 2 of this section, including the level of uncertainty for the measurement.

Clinical Information

The information presented in this section consists of examples of clinical study methodologies that the agency believes could provide valid scientific evidence to support marketing applications and should not be viewed as inflexible requirements. Marketing approval (demonstration of reasonable assurance of safety and effectiveness) could be based on successful completion of an enriched trial and comparative feature analysis. Clinical data from an enriched trial would provide an estimate of the true screening performance of the device, though performance estimates with narrower confidence intervals could only be derived from much larger data sets. The options provided in this section are based upon certain assumptions that will affect the study design and subject numbers. Manufacturers should establish their own assumptions, calculate sample sizes, and be able to justify them. Manufacturers are encouraged to consult with FDA prior to initiating the clinical studies they plan to use in support of marketing applications.

A. Trial Designs

1. Enriched Reader Study

To limit the size of the study to reasonable numbers (a “least burdensome” approach), a manufacturer could choose to evaluate standard mammographic views that included images from a population enriched with known cancers. In order to perform comparisons of sensitivity, specificity, and receiver operating characteristic (ROC) curves, ground truth about the cancer/non-cancer status of each subject must be established and known. For those women with sufficiently suspicious lesions to warrant a recommendation for (and performance of) biopsy, pathology results will constitute ground truth. For all other women, one year of follow-up without evidence of (or with development of) cancer will be accepted as ground truth.

Since these data would constitute a surrogate for screening sensitivity/specificity/ROC, device labeling concerning the screening aspect would accurately communicate this to the user. ROC curves should be constructed for both digital and analog mammography, for each trial radiologist, as well as for the pooled group of trial radiologists, based on either a discrete or quasi-continuous scaled level of suspicion of malignancy for each 4-view mammogram. Quantitative comparison between the ROC curves for digital and analog mammography requires a more elaborate analysis. Methods for obtaining confidence intervals on the difference of ROC parameters (e.g., area or partial area) between two modalities are summarized in [Appendix III](#). The confidence intervals

generated by the reader study would be reflected in labeling.

In addition to the construction of the entire ROC curves, comparisons of the sensitivities of digital and analog should be made at the BI-RADS 3-and-higher as well as at the BI-RADS 4-and-higher cut-points. For this purpose, the trial radiologists would render, in addition to a continuous scaled level of suspicion of malignancy, an action recommendation based on a BI-RADS rating.

Comparisons can be made either of specificities or of total rates of positivity, defined as biopsy recommendation. Total rate of positivity, combined with the sensitivity and the prevalence of cancer in the trial population, allows estimation of the false positive rate and of the specificity.

True and false positives and negatives should be defined at each of the two cut-points as follows:

- At BI-RADS 3-and-higher

False positives should be defined as mammograms leading to recommendation for biopsy or repeat mammogram in 6 months (or other appropriate interval shorter than the screening interval), in women either whose biopsies yield negative pathology or whose repeat mammograms fail to change within the next year, and who fail to develop clinical signs of cancer in that location within the next year.

False negatives should be defined as mammograms leading to recommendation for no further action other than return for the next screening mammogram in women who develop biopsy-proven cancer within the next year, found either through the development of clinical symptoms or a subsequent positive screening mammogram. This definition does not differentiate on the basis of whether or not the cancer is visible in retrospect or, if visible, whether there are any radiologists, blinded to subsequent outcome, who would prospectively work up the lesion; i.e., all should be included in the definition.

True negatives should be defined as mammograms leading to no further action in women who are still without evidence of cancer on their subsequent screening mammogram at least one year later.

True positives should be defined as mammograms in women with biopsy-proven malignancy leading to recommendation either a) for repeat mammogram in 6 months (approximately), which, on repeat, leads to recommendation for biopsy, or b) for immediate biopsy.

In the case of a woman who, for whatever reason, does not follow the recommendations for either 6 month follow-up or biopsy, she will be judged to have cancer, or not, depending on whether she develops clinical evidence of cancer in the same location or is still without evidence of cancer, one year from the mammogram in question.

- At BI-RADS 4-and-higher

False positives should be defined as mammograms leading to recommendation for biopsy in women with either negative pathology or (in the case of a woman who, for whatever reason, does not follow the recommendation) one year of follow-up without evidence of cancer in that location.

False negatives should be defined as mammograms leading to recommendation for either repeat mammogram in approximately 6 months, or no further action other than return for the next screening mammogram, in women who develop biopsy-proven cancer within the next year. Note that cancer may be found through either the development of clinical symptoms or a subsequent positive screening mammogram. Again the definition does not differentiate on the basis of whether or not the cancer is visible in retrospect or whether the lesion, if visible, would be worked up prospectively by any radiologist blinded as to subsequent outcome.

True negatives should be defined as mammograms leading either to recommendations for 6 month follow-up mammogram, or return for next screening mammogram, in women who are still without evidence of cancer on their subsequent screening mammogram at least one year later. Note that a true negative for the BI-RADS 4-and-higher cut-point could be a false positive for the BI-RADS 3-and-higher cut-point, and a false negative for the BI-RADS 4-and-higher cut-point could be a true positive for the BI-RADS 3-and-higher cut-point. True positives should be defined as mammograms leading to the recommendation for biopsy of a lesion, which proves to be malignant on pathology.

Women who fall into one category in one breast location and another category in another breast location will be rare enough that they can be excluded from the trial, after the fact, without significantly affecting the statistical outcome.

2. Comparative Feature Analysis Study

A comparative feature analysis is performed by radiologists examining, side-by-side, paired images of each woman obtained with two different systems, in this case an analog and a digital system. The radiologists can compare any number of different features of the images, such as conspicuity of lesions, dynamic range, etc. Such a side-by-side analysis would provide a link between a rigorous evaluation of the physical imaging performance of a system and a truth-based enriched reader study. For the comparative feature analysis, problem-solving and standard views should be used. A representative sample of the films used in this comparison should be submitted as part of the premarket application.

Whereas physical measurements cannot determine whether a digital system with a DQE inferior to that of an analog system at intermediate and high spatial frequencies is less accurate at discriminating between, say, calcifications associated with malignancy and those associated with benign processes, a feature comparison using appropriately selected features could provide useful information about this question.

- A comparative feature analysis can be conducted using either side-by-side or random-order scoring. For the volume-of-tissue-imaged feature, side-by-side ranking is appropriate. For other features, side-by-side ranking allows the introduction of reader bias, since it is unlikely that the reader can be blinded to the source (digital or analog) of any image. On the other hand, the reader's ability to distinguish small differences in feature quality is likely to be better with side-by-side interpretation. Thus use of side-by-side scoring would likely reduce the number of patients required to achieve a specified degree of discrimination with specified levels of error. The choice of scoring method is left to the manufacturer and should be accompanied by sound scientific justification.
- Either standard views, problem-solving views, or both may be used.

- Appropriate candidates for comparative features can be found in the BI-RADS™ lexicon. Other potential features of interest have been described by Sickles, Genant, and Doi (Comparison of Laboratory and Clinical Evaluations of Mammographic Screen-Film Combinations, Proc. SPIE **127**, 30-35, 1977). The feature analysis should also include an evaluation of the volume of tissue at the chest wall included in the digital image relative to that included in the image from a standard system.
- The ability to correctly identify features of malignant lesions is clearly of primary interest. This includes the ability to detect micro-calcifications, the ability to discriminate benign from malignant micro-calcifications, and the ability to discern subtle irregularities in otherwise smooth mass margins and thereby discriminate between benign and malignant masses. However, since the vast majority of screening studies will, in fact, be negative for cancer, it is also extremely important to include features from non-cancer cases that serve to distinguish between systems.
- A comparative feature analysis can be used to support the approval of changes, major and minor, to digital systems. For example, a feature analysis can provide a significant component of the basis for approval of soft-copy display subsequent to approval of hard-copy display or vice versa.

B. Presentation of Clinical Data

Manufacturers should provide all available clinical trial data in a format suitable for statistical analysis. These data should include each radiologist's Probability of Malignancy (POM) and BI-RADS categorization for each breast and each woman. The status with respect to malignancy or benignancy, based on pathology or follow-up, should be reported for each breast and each woman. Manufacturers should provide the above clinical trial data in both EXCEL and SAS data files, along with a sufficiently detailed description of each variable contained in WORD document files. In addition, manufacturers should provide all SAS programs required to convert the above raw clinical data files into ASCII data files for input into the ROC program LABMRMC (which is available from the University of Chicago Department of Radiology at <http://www-radiology.uchicago.edu/krl/toppage11.htm>). Finally, the name and e-mail address of the manufacturer's statistical consultant should be provided so that any questions concerning the clinical data can be readily addressed. The lead reviewers should be copied on such communications.

C. Post-approval Unenriched Screening Study

There are several reasons for performing postmarket studies. First, depending on the quality of the premarket data, FDA may require such studies as a condition of approval of a PMA. Second, applicants may choose to perform additional studies to address limitations in labeling, to provide additional information for prospective users, to provide additional information to third-party payers, etc.

Premarket studies could be followed by a post-approval study on an unenriched screening population, which would provide more definitive data regarding sensitivity/specificity/ROC for screening, possibly leading to revision of the labeling. In order to provide a valid estimate of the digital system's sensitivity/specificity/ROC in the screening setting, it is necessary that the case mix of mammograms, both with and without cancers, be representative of that which would be expected in a screening, rather than diagnostic/problem-solving, population. As in the premarket study, we recommend that in the post-

approval study a quasi-continuous rating of probability of malignancy (POM) be used for generating the ROC curve, and the BI-RADS categories for analysis of the patient-management decision.

The purpose of the large screening study is to estimate the effectiveness of screening mammography in actual practice. This means that readers and cases (i.e., all women being screened) should be selected as representative of the population. The individual case readings are pooled in the manner of the generic reader of paradigm C of [Appendix III](#) on ROC Analysis. The mean results will then reflect the performance of this generic reader and the uncertainties will reflect the case variability as seen by that generic reader.

Since reader-sampling variation is not measured and accounted for in the paradigm of the generic reader, a multiple-reader, multiple-case (MRMC) study according to paradigm D of [Appendix III](#) is required for this purpose. Contemporary methods for analyzing MRMC data may then be used to study the relative contributions of the sources of variability in mammographic screening (refs. 7-11 of [Appendix III](#)). A premarket study can serve as a pilot study for estimating the numbers of readers and cases required to achieve the desired precision and generalizability of estimates in the MRMC study. Note, however, that it would be impractical for multiple readers to read all of the non-cancer cases in the MRMC study. Thus, it is reasonable to randomly sub-sample the non-cancer cases, retaining only a number commensurate with the number of cancer cases (e.g., three or four times).

The post-approval non-enriched screening study could be designed such that the lower limit of the 95% confidence interval (CI) for the digital minus analog difference in ROC areas is -0.05, i.e., such that there is 95% confidence that the difference is not as negative as -0.05.

However, even if a sufficiently small difference in ROC areas (or superiority of digital over analog) is shown, if crossing ROC curves suggest the new modality might be inferior in the critical region, the non-inferiority hypothesis must also be tested specifically in that region. Also, any cancers missed by one modality should be thoroughly explored to identify systematic differences in cancer detection between digital and analog.

Additionally, sensitivity and specificity should be estimated for both digital and analog, based on patient management cut-points at BI-RADS 3 and BI-RADS 4. These sensitivity and specificity point estimates, along with corresponding 95% CI's, should be included in labeling.

Manufacturers may consider the following study design options:

- All women could be double exposed--to both analog and digital mammograms.
- Two much larger cohorts of women could be randomized, each to receive either analog or digital mammography.
- Alternatively, to avoid the selection bias inherent in double exposing only a random selection of analog negative women (along with all analog positive women), one could use the following approach: First, randomize women to either analog or digital mammography; then double expose all (analog or digital) positive women, and finally, randomly select additional women for double exposure from among those women negative on their first (analog or digital) mammogram.

Since the diagnostic ability of the new technology is a function of both the device and the staff of the mammography facility, a post-approval study design should address the range of actual users. This may

be accomplished by having each image read by radiologists with a variety of backgrounds, such as both specialists and generalists. There may be other radiologist characteristics that should also be incorporated into the strategy for their selection into the study. Since the premarket clinical study will most likely not have included such a large variety of radiologists, the increased variation in parameter estimates will need to be considered when estimating the power of the post-approval study.

Post-approval studies may be performed through participation in a large multi-device trial with a single protocol designed and performed under the auspices of a designated objective third party(ies) such as NCI or NEMA. FDA would entertain such a proposal and help to facilitate its development. Alternatively, each manufacturer could pursue its own study or other collaboration.

Hard- and Soft-copy Display

The choice of display format for the initial submission is up to the manufacturer. While hard-copy was originally understood to be the preferred medium to demonstrate agreement, in a ground truth paradigm this advantage does not apply. Therefore, manufacturers are encouraged to demonstrate the effectiveness of their devices in accordance with their marketing goals.

Marketing approval of one format will not constitute approval of the other. However, since all digital acquisitions obtained during the course of a clinical trial will be amenable to either display format, the requisite data for a subsequent evaluation of the alternative format may consist to a greater or lesser degree of comparison of the soft-copy to the hard-copy displays of the same set of mammograms used in the completed trial. In this way the extra radiation exposure of additional subjects could be minimized, if not eliminated, even though additional radiologists' readings would be required to make this comparison. (See [Appendix II](#) for further detail.)

Labeling

This section of the premarket submission provides the information associated with the proposed labeling for the digital mammography system. In a PMA, the final marketing labeling should contain the following basic elements (for guidance see: <http://www.fda.gov/cdrh/ode/labeling.pdf>).

A. Indications for Use

An essential characteristic of all labeling is that it accurately represents the data which has been collected on the device, and that it describe the disease or condition that the device will diagnose and the patient population for which the device is intended. This description should include the display mode(s) that are being marketed.

B. Contraindications

This section should list those circumstances under which the device should never be used.

C. Warnings/Precautions

When increased risk or decreased benefit is anticipated, based on available information, or is not adequately evaluated, based on study design (e.g., inclusion/exclusion criteria), such factors should be listed in this section.

D. Adverse Events

All adverse events recorded in clinical trials should be clearly presented in tabular form, including number and percent. Events should be listed in a meaningful sequence based upon incidence rates, severity, or another relevant parameter. Deaths and other significant adverse events should be described in paragraph form.

E. Summary of Non-clinical Studies.

Users of digital mammography systems should be provided with objective documentation of the imaging performance of the device. Users can employ this information in their evaluation of the importance of any trade-offs between different facets of imaging performance. This summary should include the data (as described in the Technical Data section of this guidance document) and should be detailed in the user's manual, encompassing the sensitometric response characteristics, the spatial resolution properties, the efficiency of SNR transfer, the dynamic range of the image acquisition and image display systems, the results of phantom image tests, and the patient radiation dose.

F. Summary of Clinical Studies

Clinical studies on which safety and effectiveness determinations were based should be summarized. This summary must include the 95% confidence limits associated with the various entities compared between the digital and analog systems.

G. Instructions for Use

Detailed instructions, which reflect the experience gained in preclinical and clinical studies, should be provided. This section should include, if applicable, any instructions needed as a result of using both hard-copy and soft-copy display, including any recommended settings of window levels and widths.

1. User's Manual

Provide a description of methods for selection of x-ray technique factors, image-processing algorithms, display device(s) and display parameters (e.g., window width, window level, and gray scale transfer function) in the generation of a digital mammogram. For systems using a delayed readout of image data such as a photostimulable phosphor, the total number of exposures possible on a single image receptor including the criterion for replacing the image receptor and any special precautions on fogging due to exposure to room light should be included in the manual.

2. Training Materials

A set of comprehensive training materials for digital mammography systems will help the user to move up the "learning curve" in an efficient manner. A complete description of all training and the materials to be used should be provided by the manufacturer as part of the premarket submission.

Quality Assurance Program

A submission should contain a complete description of the quality assurance program (modeled on the requirements as stated in 21 CFR 900.12 of the MQSA regulations) for the entire image acquisition and display system that includes the following information:

- A. a list of the parameters to be monitored and the frequency of monitoring;
- B. a description of the standards, criteria of quality or limits of acceptance that have been established for each of the parameters monitored;
- C. a description of the procedures to be used for monitoring each parameter;
- D. a list of the records, with sample forms (if applicable), that the facility staff must maintain to conduct the QA program; and
- E. a description of any training materials to be provided for performing, recording and monitoring quality assurance tests.

Post-Approval Requirements for Mammography Quality Standards Act (MQSA) of 1992

The following is intended to provide the manufacturer with information concerning statutory requirements imposed by MQSA some of which are similar (such as quality assurance procedures) to material normally included in a premarket application.

The facilities planning to use a full-field digital mammography (FFDM) system must be informed that the FFDM system must be certified under MQSA before non-investigational clinical examinations can be performed with the system. The requirements of 21 CFR 900.12 of the MQSA regulations establish the minimum quality standards that must be met by a facility to be eligible for certification to provide screening and/or diagnostic mammography services. These requirements include standards for personnel, equipment characteristics, a quality assurance program for equipment, and reporting and record keeping requirements. While some of these standards focus on screen-film equipment, many of them apply to all mammography facilities. There are also additional standards that apply to the use of FFDM systems. The manufacturers of FFDM systems should assist the facilities in meeting the following additional requirements to become certified for use of their systems under MQSA.

Before a digital mammography system can be used for non-investigational examinations, the facilities must establish a quality assurance program recommended by the digital mammography system manufacturer. This program must include quality control (QC) test procedures with appropriate test frequencies. The QC manual describing the program must be considered as a regulatory document and must include proper control and or action limits for the QC tests. The action limit for the maximum allowable mean glandular dose shall not exceed the maximum allowable dose for screen-film mammography. When the test results fall outside of the action limits, corrective actions must be successfully carried out within appropriate time frames before the equipment can again be used for clinical examinations.

If hard-copy images are used for clinical or phantom images, the QC manual must include a detailed description of the printer(s) with QC test criteria and test frequency with proper action limits. If the soft-copy display is used for image displays, detailed descriptions of display monitor requirements with QC test procedures with appropriate frequencies and action limits must be included in the QC manual.

All personnel performing mammography services with the digital mammography systems must have appropriate modality-specific training prior to using the system, as well as meeting the general personnel requirements.

The system must be evaluated by a MQSA qualified medical physicist before clinical examinations can be performed with the system.

Appendix I

Definitions of Terms

Since there are ambiguities in a number of terms concerning mammography, the following is a set of definitions used in this document:

Symptomatic: Having a clinically (as opposed to mammographically) detectable breast abnormality, such as a palpable mass, nipple discharge or nipple inversion, etc. For the purposes of this document the definition of symptomatic does not differentiate between signs and symptoms, but rather includes both.

Asymptomatic: Having no clinically detectable signs or symptoms. In particular, asymptomatic women with abnormal screening mammograms are still considered to be asymptomatic.

Screening mammogram: Standard 4-view mammogram, i.e., medio-lateral oblique (MLO) and cranio-caudal (CC) views of each breast, performed on asymptomatic women to detect the presence of breast cancer. Repeat standard views done for technical reasons, such as inadequate compression, inadequate amount of included breast tissue, or over- or underexposure, are included in this definition.

Short-term follow-up mammogram: Repeat standard-view mammogram of either one or both breasts on an asymptomatic woman whose screening mammogram was assigned to the BI-RADS 3 category. This is generally done within 4 to 6 months of the screening mammogram, but always within less than a year.

Diagnostic mammogram: Following general use in clinical practice, this term includes not only standard-view mammograms on symptomatic women, as well as standard-view short-term follow-up mammograms on asymptomatic women, but also all special/problem-solving views, whether on symptomatic women needing views in addition to a standard-view mammogram or on asymptomatic women needing views in addition to a standard-view screening or short-term follow-up mammogram (see Table below).

Special/problem-solving views: Mammographic views other than standard MLO or CC--such as rolled, spot compression, magnification, tangential, latero-medial, pinched or implant-displaced, etc.--performed for non-technical reasons, either on asymptomatic women with incomplete (i.e., BI-RADS 0) screening or short-term follow-up mammograms, or on symptomatic women with incomplete standard-view mammograms. In particular, in accordance with the above, repeat standard views performed for technical reasons, are excluded from this definition.

Table. Shaded area represents diagnostic mammography.

Views	Patients	Symptomatic	Asymptomatic	
			Screening	Short-term follow-up
	Standard			
	Special/ Problem Solving			

Appendix II

Approval of Soft-copy (Hard-copy) Display Subsequent to Hard-copy (Soft-copy)

Non-Clinical Information

This part of the guidance describes the non-clinical content of marketing applications for the approval of image display on a soft-copy device subsequent to the approval of image display on a hard-copy device or vice versa for a digital mammography system. This appendix addresses the differences in the physical characteristics and the image processing associated with the two display devices. There are additional information requests associated with the display device, whether soft-copy or hard-copy throughout the main body of the guidance document. The device characteristics should include the following:

- a complete description of the differences in functional components associated with the two display devices including a pictorial representation of the layout and interconnection of the different components
- a complete description of the differences in the grayscale display function associated with the two display devices, e.g., perceptual linearization

Clinical Information

This section presents examples of study protocols that could provide the basis for approval of soft-copy (hard-copy) display subsequent to approval of hard-copy (soft-copy) display. A clinical study should be designed to supplement the evaluation of the physical parameters of a soft-copy (hard-copy) device with clinical data.

Note: A representative sample of the films used in the clinical study should be submitted as part of the premarket application.

A. Reader Study

One approach is to use a truth-based reader study design similar to that originally used for the premarket approval of the digital mammography system. In this case, the comparison would be between the display of the **same** digital image on a soft-copy (hard-copy) display device versus on a hard-copy (soft-copy) device. As such, the clinical study would not be subject to some of the variability associated with the physical positioning of the breast. It should be noted, however, that concerns regarding intra- and inter-reader variability would still apply.

B. Comparative Feature Analysis Study

Another approach would be use of the comparative feature analysis model using problem-solving and standard views. This approach is attractive since it may lend itself to the comparison between display systems with a much smaller number of images than is required for a truth-based reader study. The same caveats described under Comparative Feature Analysis Study (Clinical Information section A2, p.12-13) would apply.

Labeling

A. Summary of Non-clinical Studies

The labeling should include a summary of the data as requested in the Non-clinical Information section of the guidance.

B. Summary of Clinical Studies

The labeling should include a summary of the data resulting from the clinical study as requested in the Clinical Information section of the guidance.

C. Instructions for Use

Detailed instructions and training materials which reflect the experience gained in non-clinical and clinical studies should be provided, including any instructions needed, as a result of using the soft-copy device, to fulfill the indications statement.

Anticipated MQSA Requirements for Soft-copy (Hard-copy) Devices

Relevant information includes a complete description of the quality assurance program for the soft-copy (hard-copy) device as follows:

- a description of the standards, criteria of quality or limits of acceptance that have been established for each of the parameters monitored
- a description of the procedures to be used for monitoring each parameter
- a list of the records, with sample forms (if applicable), that the facility staff must maintain to conduct the QA program
- a description of any training materials to be provided for performing, recording and monitoring quality assurance tests

Appendix III

ROC Methodology

The receiver (or relative) operating characteristic (ROC) curve is the plot of the true-positive fraction (TPF) vs. the false-positive fraction (FPF), both of which vary with the level of aggressiveness of the reader of a diagnostic test. In imaging studies an entire ROC curve is usually measured on a single pass from the graded responses provided by an image reader for a sample of cases drawn from a specified population (refs. 1, 2). Until recent years a five-category rating scale, or level of suspicion of disease, was commonly used. A number of contemporary investigators (ref. 3 and their followers) favor the use of a quasi-continuous 100-point probability-rating scale as a more natural approach. Others prefer a version of this that retains ten bins, namely, 0-10%, 11-20%, . . . , 91-100% probability of disease, here, breast cancer. (Some investigators in mammography use 0-2%, 2-20%, 20-30%, etc.) Software is available over the Internet for all of these approaches (refs. 4, 5).

Over and above the probability-rating report used for ROC analysis, the ACR BI-RADS scale is used for reporting and assessing the patient-management decision in studies of mammography. This has already been referred to earlier. (Note that the inclusion of the probability rating that is required for ROC analysis is also within the spirit of the BI-RADS recommendations, in particular, the annotation under Category 4 in that document.)

Analysis of uncertainties in estimates of ROC parameters and differences in these parameters between imaging modalities is an essential part of a submission of a clinical study. The particular approach to uncertainty analysis selected by a manufacturer will determine the level of the effectiveness claim and labeling that can be considered for approval. There are several experimental and analytical paradigms that have been used for uncertainty analysis in ROC studies (refs. 6 – 11):

A. The simplest paradigm is that of a single reader reading a sample of cases from a specified population. Estimates of mean performance of a modality and mean differences between modalities, together with uncertainties derived from measurements on a single reader and a finite case sample, are said to be “generalizable to the population of cases and that particular reader.”

B. An elaboration of the previous paradigm is to study a number of readers, each independently reading the same cases, and to present separately the individual ROCs for all of the readers. Although rating-data sets from such studies may be pooled or the resulting curves combined, there is no method for obtaining a meaningful estimate of uncertainty from such a pooling or combination. However, the individual-reader ROC curves for a given diagnostic modality and the corresponding results for a second modality obtained in this way are subject to tests of the significance of the difference on a reader-by-reader basis, and reporting of reader-by-reader results. As in paradigm A above, the results for each reader are only “generalizable to the population of cases and *that particular reader*.” An alternative approach is to test the significance of the difference of ROC summary measures (e.g., ROC area or partial area) between two modalities based on a list of such individual-reader sample results; in this case the result is said to be “generalizable to a population of readers and *the particular case sample* of the study.” Case-sample variation is not accounted for in this alternative approach.

C. A different approach has been to combine all image *probability ratings* from a set of readers or institutions, each independently reading their own local collection of images just once, into a pool that represents the performance of a “generic reader” (ref. 12). One may then use traditional approaches to

estimate the uncertainty due to an effective case variability as seen by that generic reader. This uncertainty, however, is generalizable neither to a population of generic readers (since only one such reader has been sampled) nor to a population of readers in general (since reader-sampling variation is not accounted for in such an analysis). Nevertheless, such an approach may be a practical expedient for very large studies, including the possibility of a number of generic readers reading the same cases.

D. Contemporary approaches to ROC methodology include the multiple-reader, multiple-case (MRMC) paradigm (or “reader study”), in which every reader reads every case, and random-effects models are used to account for both case variability and reader variability (refs. 7 – 11). The MRMC approach requires elaborate analysis of the reported levels of suspicion of disease rather than pooling of data or averaging of ROC curves. ROC parameter estimates and uncertainties derived from this approach are said to be “generalizable to the population of cases and population of readers” that were sampled in the study. When two modalities are compared, statistical power is gained when the same patients are used for both modalities, and when the same (or well-matched) readers are used for both modalities, in both cases due to the correlation that is introduced across modalities. (When this commonality or matching is absent, it is necessary to consider so-called “split-plot” designs and analysis (ref. 11).)

REFERENCES

1. Swets JA, Pickett RM. Evaluation Of Diagnostic Systems: Methods From Signal Detection Theory. New York, NY: Academic Press, 1982.
2. Metz CE. ROC Methodology In Radiologic Imaging. Invest Radiol 1986; 21: 720-33.
3. Rockette HE, Gur D, and Metz CE. The Use Of Continuous And Discrete Confidence Judgements In Receiver Operating Characteristic Studies Of Diagnostic Imaging Techniques. Invest Radiol 1992; 27: 169-172.
4. Web address for University of Chicago software:
<http://www-radiology.uchicago.edu/krl/toppage11.htm#software>
5. Web address for University of Iowa software: <ftp://perception.radiology.uiowa.edu>
6. Metz CE. Some Practical Issues Of Experimental Design And Data Analysis In Radiological ROC Studies. Invest Radiol 1989; 24: 234-245.
7. Dorfman DD, Berbaum KS, Metz CE. Receiver Operating Characteristic Rating Analysis: Generalization To The Population Of Readers And Patients With The Jackknife Method. Invest Radiol 1992; 27: 723-731.
8. Gatsonis CA, Begg CB, Wieand S. Advances in Statistical Methods for Diagnostic Radiology: A Symposium. Acad Radiol 1995; 2 (Supplement 1, entire issue, esp. papers by CA Beam, CA Gatsonis, NA Obuchowski, CA Gatsonis and A Toledano).
9. Toledano A, and Gatsonis C. Ordinal Regression Methodology For ROC Curves Derived From Correlated Data. Statistics in Medicine 1996; 15: 1807-1826.
10. Beiden SV, Wagner RF, Campbell G. Components-Of-Variance Models And Multiple-Bootstrap Experiments: An Alternative Methodology For Random-Effects ROC Analysis. Acad Radiol 7; 341-349 (2000)
11. Dorfman DD, Berbaum KS, Lenth RV, Chen Y-F. Monte Carlo Validation Of A Multireader Method For Receiver Operating Characteristic Discrete Rating Data: Split Plot Experimental Design. Proc. of the SPIE 1999; 3663: 91-99 (Bellingham WA).
12. Baker SG and Pinsky PF. A Proposed Design and Analysis for Comparing Digital and Analog Mammography: Special ROC Methods for Cancer Screening. JASA 2001 (In Press).