A Multi-Reader Multi-Case Controlled Clinical Trial to Evaluate the Comparative Accuracy of Fujifilm DBT plus S-View versus FFDM Alone in the Detection of Breast Cancer

A Pivotal Study

FUJIFILM Medical Systems U.S.A., Inc.

419 West Avenue

Stamford, CT 06902

Table of Contents

Abbreviations

Abbreviation	Explanation
ACR®	American College of Radiology
AE	Adverse Event
ANOVA	Analysis of Variance
AUC	Area Under the Receiver Operating Characteristic (ROC) Curve
BI-RADS®	Breast Imaging Reporting and Data System
CC	Cranio-Caudal
CFR	Code of Federal Regulations
CI	Confidence Interval
DBT	Digital Breast Tomosynthesis
DCF	Data Clarification Form
DMC	Data Management Center
eCRF	Electronic Case Report Form
eDC	Electronic Data Capture
FP	False Positive
FDA	Food and Drug Administration
FFDM	Full-Field Digital Mammography (also known as Digital Mammography)
GCP	Good Clinical Practice
HIPAA	Health Insurance Portability and Accountability Act
ICH	International Conference on Harmonization
ID	Identification
IDC	Invasive Ductal Carcinoma
IDE	Investigational Device Exemption
ILC	Invasive Lobular Carcinoma
IRB	Institutional Review Board

LCC	Left Cranio-Caudal
LMLO	Left Medio-Lateral Oblique
MDR	Medical Device Reporting
MLO	Medio-Lateral Oblique
MQSA	Mammography Quality Standards Act
MRMC	Multi-Reader Multi-Case
PHI	Protected Health Information
PMA	Premarketing Approval Application
POM	Probability of Malignancy
RCC	Right Cranio-Caudal
RMLO	Right Medio-Lateral Oblique
ROC	Receiver Operating Characteristic
S-View	Synthesized View
STARD	Standards for Reporting of Diagnostic Accuracy
TP	True Positive

Definitions

ACR® BI-RADS® Breast Tissue Composition (Density) Categories	a. The breasts are almost entirely fatty b. There are scattered areas of fibroglandular density c. The breasts are heterogeneously dense, which may obscure small masses d. The breasts are extremely dense, which lowers the sensitivity of mammography
ACR® BI-RADS® Assessment Categories	Assigned by reader; **0** = Incomplete – Need Additional Imaging Evaluation and/or Prior Mammograms for Comparison **1** = Negative **2** = Benign **3** = Probably Benign. Management: Short-interval (6-month) follow-up or continued surveillance mammography **4** = Suspicious. Management: Tissue diagnosis. **4A** = Low suspicion for malignancy **4B** = Moderate suspicion for malignancy **4C** = High suspicion for malignancy **5** = Highly Suggestive of Malignancy
BI-RADS® Score	Derived variable based on lesion matching and requiring correct lesion localization
ASPIRE Bellus II	Fujifilm Mammography Viewer (see Appendix A) where the FFDM and DBT plus S-View images are displayed and reviewed by the physician.
ASPIRE Cristalle with DBT Option	Fujifilm's Mammography Acquisition System (see Appendix A) which generates both FFDM and DBT images.
ASPIRE Cristalle DBT Option with S-View	ASPIRE Cristalle DBT Option (P160031) with synthesized view (S-View), where the S-View images are created from the reconstructed DBT images. "CAUTION: Investigational Device Limited by U.S. Federal Law to Investigational Use."
Benign Case	A case in which all lesions are confirmed as benign by a biopsy or surgery and no other lesions are biopsy-malignant.
Cancer Case	A case in which at least one lesion is confirmed as malignant by biopsy or surgery.
Cancer Recall Rate	The rate (or fraction) of cancer case(s) correctly recalled by the reader
Correct Lesion Localization	The location and type of a finding annotated by the reader matches those of a lesion annotated by the truther, as determined by the Lesion Matcher.
DBT Examination	Each DBT Examination has two parts: • Acquisition: Acquisition of DBT images for each of the standard bilateral CC and MLO views (RCC, LCC, RMLO, LMLO) using the ASPIRE Cristalle. • Review: DBT images are displayed and interpreted by the physician on the ASPIRE Bellus II.
DBT Images	A series of mammograms, which are reconstructed from multiple low-dose projection images captured by the ASPIRE Cristalle with the x-ray tube rotated at a number of small offset angles over a limited angular range while the breast remains compressed and unmoved.

DBT Read	The physician reads (interprets) the DBT images displayed on the ASPIRE Bellus II (see Appendix A)
DBT Views	A series of cross-sectional views of the DBT images for each of the CC and MLO views for each breast.
DBT plus S-View Examination	Each DBT plus S-View Examination has two parts: • Acquisition: Acquisition of DBT images for each of the standard bilateral CC and MLO views (RCC, LCC, RMLO, LMLO) using the ASPIRE Cristalle. • Review: DBT plus S-View images are displayed and interpreted by the physician on the ASPIRE Bellus II.
DBT plus S-View Images	DBT plus S-View images for each of the RCC, LCC, RMLO, LMLO for each breast.
DBT plus S-View Read	The physician reads (interprets) the DBT plus S-View images displayed on the ASPIRE Bellus II (see Appendix A).
Evaluable Subject	Based on acquisition-site data, an evaluable subject is a subject with known true clinical status and with complete DBT and FFDM examinations (four standard views), in which there is sufficient anatomical coverage, sufficient contrast, no significant motion or other artifacts, limited noise, and similar positioning.
False Positive	A false positive is when a negative case has been incorrectly identified as a positive case.
FFDM Examination	Each standard FFDM Examination has two parts: • Acquisition: Acquisition of FFDM images for each of the standard bilateral CC and MLO views (RCC, LCC, RMLO, LMLO). • Review: FFDM images are displayed and interpreted by the physician on the Fujifilm Mammography Viewer.
FFDM Images	Standard FFDM mammograms; comprised for the four standard views (RCC, LCC, RMLO and LMLO) images captured by the ASPIRE Cristalle (see Appendix A).
FFDM Read	The physician interprets the FFDM images displayed on the Fujifilm Mammography Viewer (see Appendix A).
FFDM Views	Standard CC and MLO views for each breast.
FMSU2013-004A	Acquisition of Digital Mammography and Breast Tomosynthesis Images for Clinical Evaluation of Fujifilm Digital Breast Tomosynthesis
FMSU2017-002A	A Multi-Reader Multi-Case Controlled Clinical Trial to Evaluate the Comparative Accuracy of Fujifilm DBT plus S-View versus FFDM Alone in the Detection of Breast Cancer – A Pilot Study
FMSU2017-002B	A Multi-Reader Multi-Case Controlled Clinical Trial to Evaluate the Comparative Accuracy of Fujifilm DBT plus S-View versus FFDM Alone in the Detection of Breast Cancer – A Pivotal Study
Forced BI-RADS	A BI-RADS assessment category of 1 through 5, which the readers must give if they had given the case an initial BI-RADs score of 0. This score is based solely on the mammogram images presented to the reader and not the actual results of any additional work-up that may have been performed pertaining to that specific case.
Fujifilm	FUJIFILM Medical Systems U.S.A., Inc.
IA Site DBT Reader	An image-acquisition site reader is a qualified radiologist who interprets the DBT

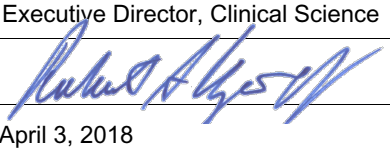
	images at the time of image acquisition (see FMSU2013-004A) and, in doing so, contributes to the reference standard for those images.
IA Site FFDM Reader	An image-acquisition site reader is a qualified radiologist who interprets the FFDM images at the time of image acquisition (see FMSU2013-004A) and, in doing so, contributes to the reference standard for those images.
Initial BI-RADS	In the reader study, a BI-RADS assessment category of either a 0 (incomplete), 1 (negative) or 2 (benign); used as a preliminary assessment of a case.
Lesion Location	The location of the lesion marked by the reader study reader.
Lesion Matcher	An expert not associated with diagnosing cases at the image-acquisition sites or serving as a reader interpreting images on this reader study. This expert may be an expert mammography technologist with an active license or a board-certified radiologist and MQSA-qualified mammographer; if a board-certified radiologist and MQSA-qualified mammographer, this expert may be the same person as the truther.
Mammographically Detectable Finding	A mammographically detectable finding is a breast finding detectable by either FFDM or DBT plus S-View as determined by the truther.
Negative Case	A negative case is defined as a case where there are no abnormal areas or malignant (cancer) findings within a year (within 320 to 455 days) from the original screening examination.
Non-Cancer Case	A biopsy benign case, or a recall case, or a normal case.
Non-Cancer Recall Rate	The rate (or fraction) of non-cancer cases incorrectly recalled by the reader.
Per-Subject Analysis Requiring Correct Localization	Per-subject analysis is the analysis of BI-RADS®, POM, or recall ratings assigned to the study subject requiring correct lesion localization.
POM Score	The study reader's assigned probability that a finding is malignant (POM score ranges from 0 to 100%).
Prior Mammogram	Historical mammogram acquired on any media – including screen-film, digitized image, FFDM and/or DBT.
Recall Case	A case where there is a suspicious area(s) identified in the screening exam that needs additional work-up to determine if the area is cancerous or not. This is determined at the IA site.
Recall Score	The recall score indicates whether a finding will be recalled or not (e.g., recall score "1" is recall and "0" is no-recall).
Reference Standard	The reference standard (also often called the "gold standard" or "ground truth" in the imaging community) for patient data indicates whether the disease/condition/abnormality is present and may include such attributes as the extent or location of the disease/condition/abnormality.
Reference Standard for Cancer Cases	Biopsy or surgery: At least one lesion is confirmed as malignant by biopsy or surgery.
Reference Standard for Non-cancer Cases	Established by biopsy or surgery for benign cases, or interpretation by radiologists at the enrolling site for recall and normal cases. **Note**: Under the FMSU2013-004A image acquisition protocol, subjects with no lesions biopsied were to have a 1-year follow-up mammogram (collected within 320 to 455

	days following the original mammogram) with no malignancy discovered.
ROC Curve	A Receiver Operating Characteristic (ROC) curve is a plot of Sensitivity versus 1-Specificity and is a summary of diagnostic performance of a device or a clinician.
Screening Population	Asymptomatic subjects who present for screening mammograms (FFDM and/or DBT).
Sensitivity	Sensitivity is defined as the probability that a test is positive for a population of patients with the disease/condition/abnormality.
Specificity	Specificity is defined as the probability that the test is negative for a population of patients without the disease/condition/abnormality (i.e., normal subjects).
Standard views	Standard CC and MLO views for each breast (RCC, LCC, RMLO, LMLO).
Subject reference standard	The standard against which blinded readers' interpretations will be compared to determine the accuracy of their interpretations.
Symptomatic Subjects	Subjects recommended for diagnostic digital mammography and/or breast tomosynthesis based on screening mammogram and/or breast tomosynthesis results or clinically detectable signs and symptoms.
Synthesized View (S-View) Images	The synthesized view image for each of the LCC, RCC, LMLO, or RMLO views is created from "combining" the DBT images of that view to a single synthetic view image.
True Positive	A true positive is when a positive case has been correctly identified as a positive case.
Truther	An expert radiologist not associated with diagnosing cases at the image-acquisition sites or interpreting images in the reader study who will determine reference standard status and annotate lesions.
Truthing Process	The truthing process describes the process that the truther follows to confirm the reference standard for each case, based on materials from the image acquisition site including the study subject's FFDM and DBT plus S-View images. For each case with at least one (1) cancerous lesion, the truther will review the images and identify, confirm, and annotate lesions (including the types of lesions visible within the FFDM and/or DBT plus S-View images, the locations of the lesions, and the characteristics of the lesions) on the eCRF.

Tables and Figures

1. Sponsor Protocol Approval Signature Page

Study Title	A Multi-Reader Multi-Case Controlled Clinical Trial to Evaluate the Comparative Accuracy of Fujifilm DBT plus S-View versus FFDM Alone in the Detection of Breast Cancer - A Pivotal Study
Protocol Number	FMSU2017-002B
Effective Date	March 13, 2018
Study Sponsor	FUJIFILM Medical Systems U.S.A., Inc. 419 West Avenue Stamford, Connecticut 06902, U.S.A.

FUJIFILM Medical Systems U.S.A., Inc.	**Approval**
Name, Title	Robert A. Uzenoff Executive Director, Clinical Science
Signature	
Date	April 3, 2018

2. Investigator Agreement Signature Page

Study Title	A Multi-Reader Multi-Case Controlled Clinical Trial to Evaluate the Comparative Accuracy of Fujifilm DBT plus S-View versus FFDM Alone in the Detection of Breast Cancer - A Pivotal Study
Protocol Number	FMSU2017-002B
Effective Date	March 13, 2018
Study Sponsor	FUJIFILM Medical Systems U.S.A., Inc. 419 West Avenue Stamford, Connecticut 06902, U.S.A.
Principal Investigator	Laurie L. Fajardo, M.D., MBA, FACR, FSBI Radiology Consultant 1730 Lucky John Dr. Park City, UT 84060

I, the principal investigator named above, acknowledge the receipt of this protocol. I have read this protocol in its entirety and agree to conduct this study according to this protocol and all applicable regulations/guidelines.

Principal Investigator	Laurie L Fajardo, MD, MBA, FACR, FSBI
Signature	Laurie L Fajardo, MD
Date	April 3, 2018

3. Protocol Synopsis

SPONSOR NAME	FUJIFILM Medical Systems U.S.A., Inc.
PROTOCOL NUMBER	FMSU2017-002B
STUDY TITLE	A Multi-Reader Multi-Case Controlled Clinical Trial to Evaluate the Comparative Accuracy of Fujifilm DBT plus S-View versus FFDM Alone in the Detection of Breast Cancer – A Pivotal Study
DEVICE NAME	Fujifilm ASPIRE Cristalle with DBT Option
PURPOSE	The purpose of the pivotal reader study is to assess the comparative accuracy of Fujifilm DBT plus S-View versus FFDM in the detection of breast cancer. Radiologists' performance metrics for the following modalities will be evaluated: 1. Standard FFDM read on the ASPIRE Bellus II workstation ("**FFDM**") 2. Synthesized (S-View) read in conjunction with DBT read on the ASPIRE Bellus II workstation ("**DBT plus S-View**") This retrospective, multi-reader multi-case (MRMC) study will have an enriched sample of approximately 300 cases: 60 cancer cases and 240 non-cancer cases (including up to 48 biopsy proven benign cases and 72 recall cases). Enrichment is with respect to proportions of cancer cases, biopsy benign cases, and recall cases. As its primary endpoint, this study is designed to evaluate whether the area under the receiver operating characteristic (ROC) curve (AUC) based on probability of malignancy (POM) scores and requiring correct lesion localization is statistically significantly non-inferior for DBT plus S-View versus FFDM. Multiple secondary endpoints are outlined in the next section.
STUDY ENDPOINTS	The safety and effectiveness of DBT plus S-View are both linked to the ability of radiologists interpreting DBT plus S-View images to accurately diagnose breast cancer. Each study endpoint therefore addresses both safety and effectiveness. All of the following endpoints are measured on the FFDM read, DBT plus S-View read, and their difference. All measurements require correct lesion localization. The **primary endpoint** is: • Non-inferior per-subject average area under the ROC curve (AUC) for DBT plus S-View versus FFDM, based on POM scores and requiring correct lesion localization The study will be considered to have successfully demonstrated safety and effectiveness of the Fujifilm ASPIRE Cristalle DBT plus S-View system if the per-subject average AUC for DBT plus S-View is statistically significantly non-inferior to the average AUC for FFDM at the alpha = 0.05 significance level, for non-inferiority margin delta = 0.05. This will be established if the lower limit of the two-sided 95% CI for the difference in average AUC for DBT plus S-View – FFDM lies entirely above –0.05.

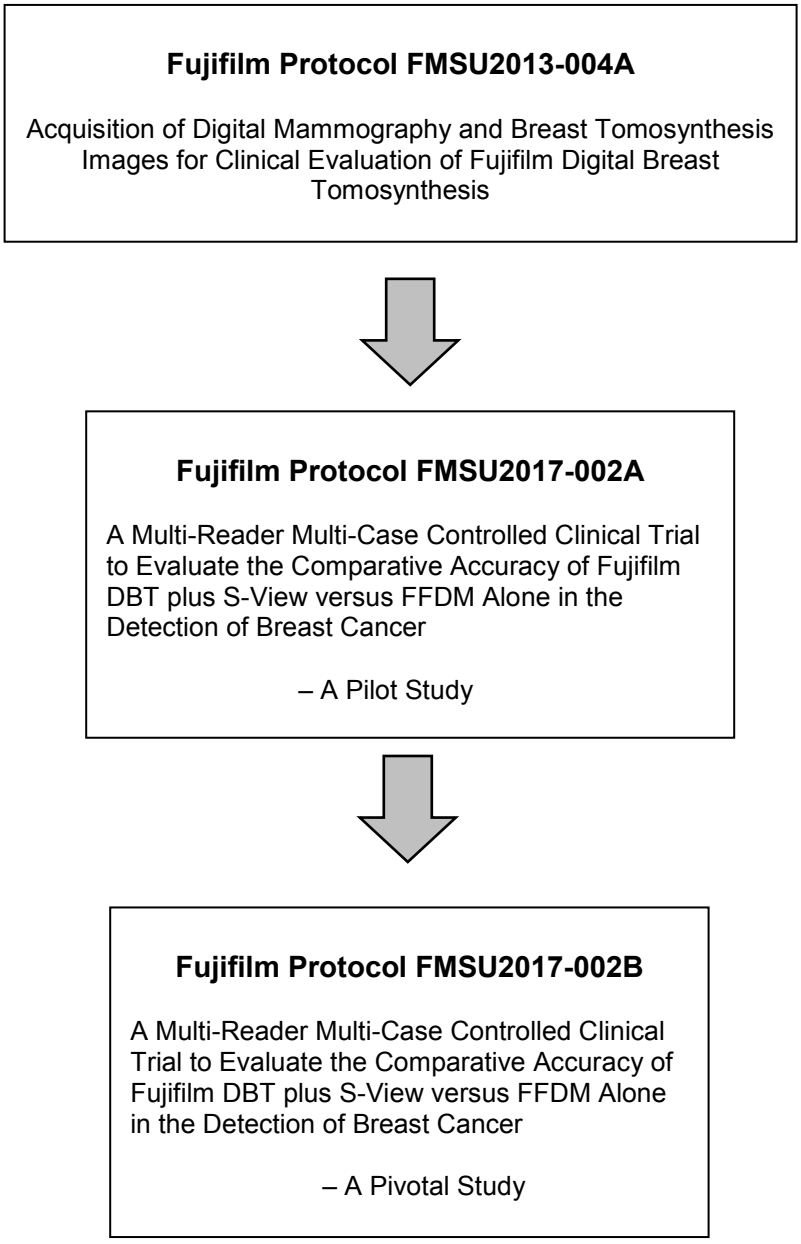
	The **secondary endpoints** are**:** 1. Non-inferior and/or superior (lower) per-subject average recall rate for all non-cancer cases for DBT plus S-View versus FFDM, based on recall scores, using non-inferiority margin delta = 0.05. 2. Non-inferior per-subject average recall rate for DBT plus S-View versus FFDM for all cancer cases, based on recall score, using non-inferiority margin delta = 0.10 3. Non-inferior and/or superior per-subject average sensitivity for DBT plus S-View versus FFDM, based on BI-RADS scores requiring correct lesion localization, using non-inferiority margin delta = 0.10 4. Superior per-subject average AUC for DBT plus S-View versus FFDM, based on POM scores and requiring correct lesion localization 5. Non-inferior and/or superior per-subject average specificity for DBT plus S-View versus FFDM, based on BI-RADS scores 6. Superior (lower) per-subject average recall rate for DBT plus S-View versus FFDM for all follow-up proven non-cancer recall cases 7. Non-inferior per-lesion average sensitivity for masses, masses with calcifications, focal asymmetries, and/or architectural distortions for DBT plus S-View versus FFDM, based on BI-RADS scores requiring correct lesion localization, using non-inferiority margin delta = 0.10 8. Non-inferior and/or superior per-lesion average sensitivity for calcifications for DBT plus S-View versus FFDM, based on BI-RADS scores requiring correct lesion localization, using non-inferiority margin delta = 0.10 9. Non-inferior (margin delta = 0.05 for AUC, 0.10 for other performance metrics) and/or superior average AUC and/or other performance metric(s) for DBT plus S-View versus FFDM for subjects with dense breasts (BI-RADS breast composition categories c. The breasts are heterogeneously dense, which may obscure small masses and d. The breasts are extremely dense, which lowers the sensitivity of mammography)
CORE READING CENTER	The pivotal study will be performed at International HealthCare, LLC in Norwalk, CT.
PRINCIPAL INVESTIGATOR, TRUTHER, READERS, AND LESION MATCHER	The principal investigator for this study is Laurie Fajardo, MD, MBA, FACR, FSBI. The truther, the lesion matcher, and the pivotal study radiologist readers shall meet all criteria for their respective roles as specified in the investigational plan. The truther, the lesion matcher, and the pivotal study readers will, respectively, participate in a training session to ensure their understanding of the ASPIRE Bellus II workstation, the protocol, the study eCRFs, study processes, GCP, and their obligations as study participants.
PIVOTAL READER STUDY POPULATION - **SAMPLE SIZE AND CASE MIX** - **CASE SELECTION AND TARGETED DISTRIBUTION**	This clinical research is a retrospective, pivotal, multi-reader, multi-case (MRMC) study with an enriched sample of 300 breast screening or diagnostic cases which will be selected from the library of mammograms collected under Fujifilm protocol FMSU2013-004A. All subjects in the FMSU2013-004A library previously provided written informed consent. As part of the consent process, subjects agreed that image data and supporting documentation could be used for future research and investigations. Evaluable cases are defined as subjects who met all protocol inclusion/exclusion

	criteria (see Appendix B), whose true clinical status is known, and who have a complete set of adequate-quality mammography examinations. Image quality is considered adequate (as determined by the image-acquisition sites under the protocol) when there is sufficient anatomical coverage and contrast, and when there is no significant effect from motion, artifacts, or noise. Measures to ensure subject and reader confidentiality include assignment of study and reader unique ID numbers. 1. Sample Size and Case Mix Enriched sample size of approximately 300 cases (approximately 60 cancer cases and approximately 240 non-cancer cases) with the following case mix: • Approximately 60 cancer cases, 48 biopsy benign cases, 72 recall cases, and 120 normal cases. 2. Case Selection and Targeted Distribution The statistician will select cases for this pivotal reader study and cases will be randomly selected to provide representative distributions of breast composition, finding types, and equal distribution across enrolling sites, to the extent that this can be accomplished using cases in the FMSU2013-004A case library.
STUDY DESIGN - **CASE REFERENCE STANDARDS** - **THE TRUTHING PROCESS** - **THE READING PROCESS** - **THE LESION MATCHING PROCESS**	This clinical research is a retrospective, pivotal, multi-reader, multi-case (MRMC) study with an enriched sample of approximately 60 cancer cases and approximately 240 non-cancer cases. 1. Case Reference Standards • The reference standard for a cancer case is a case in which at least one lesion is confirmed as malignant by biopsy or surgery. • The reference standard for a non-cancer case is biopsy or surgery for benign cases, or interpretation by radiologists at the acquisition site for recall and normal cases.
	1. The Truthing Process The truther will review all cancer cases selected for inclusion in the reader study. The truther will complete an eCRF for each case, based on materials from the image acquisition site including mammograms. For each subject with at least one (1) biopsy proven malignant lesion, the truther will review the images and identify, confirm, and annotate lesions including the types of lesions visible within the mammogram and DBT image, the locations of the lesions, and the characteristics of the lesions on the eCRF.

	2. The Reading Process Approximately 18 qualified radiologists will independently perform two reads on all (approximately 300) cases. Each reader will read each case both as a FFDM read and a DBT plus S-View read, all on the ASPIRE Bellus II workstation. Each reader will read half the cases on FFDM and the other half of the cases on DBT plus S-View during one visit. The reader will then read the complementary FFDM and DBT plus S-View during the second visit. In order to minimize memory recall bias, the two reading visits will be separated by a memory washout period of approximately four weeks. Interpretation instructions: The reader will complete the eCRFs for each case for the two reads. For each case on each read, the reader will first record whether there are mammographic findings and an initial BI-RADS assessment of 0, 1, or 2. If the answer to this question is "no" the reader will be asked to confirm for this case: a BI-RADS assessment category of 1 or 2, a POM score following the POM guidance provided in Table 4, and recall decision of "no". If the reader answers "yes" to whether there are mammographic findings, the reader will be asked to confirm an initial BI-RADS assessment category of 0, and will then provide detailed information on each suspicious finding (up to three findings) and the decision whether or not to recall the subject, as well as overall POM score and BI-RADS assessment category.
	3. The Lesion Matching Process An expert not associated with diagnosing cases at the image-acquisition sites or serving as a reader interpreting images on this pilot reader study will conduct lesion matching. This expert may be an expert mammography technologist with an active license or a board-certified radiologist who is MQSA-qualified for both FFDM and tomosynthesis; if a board-certified radiologist who is MQSA-qualified for both FFDM and tomosynthesis, this expert may be the same person as the truther.
STUDY ANALYSIS	All of the primary, secondary, and additional (optional) endpoints are measured on the FFDM read and DBT plus S-View read, and their difference. All measurements require correct lesion localization. 1.1. Analysis of primary endpoint The primary endpoint on this study is per-subject AUC requiring correct lesion localization. We will estimate AUCs for each reader in each review condition (FFDM, DBT plus S-View) based on per-subject POM scores. We plan to perform MRMC comparison of AUCs between FFDM and DBT plus S-View using the standard MRMC analysis of variance (ANOVA) method of Obuchowski and Rockette (1995)[11] to ensure generalization of the study results both to the population of readers and the population of cases. Two-sided 95% confidence intervals (CIs) will be used to quantify uncertainty in the within-modality estimates and the between-modalities difference. 1.2. Analysis of secondary endpoint(s) Analysis of per-subject recall rates, specificity, and/or sensitivity will use standard MRMC ANOVA methods. Analysis of per-lesion sensitivity will use methods for clustered data from MRMC studies that take into account the correlation between

	lesions in the same case. Subgroup analyses in non-cancer case types, lesion types, and/or breast density cohorts will use similar methods. 1.3. Multiple Comparisons To protect the study's type I error rate from inflation we will use the iterative graphical approach described in Bretz, et al. (2009)[13] to sequentially reject hypotheses.

4. Fujifilm FMSU2013/2017 Series Protocols



5. Investigators and Study Administrative Structure

Study Personnel

Principal Investigator

Laurie L. Fajardo, MD, MBA, FACR, FSBI
Radiology Consultant
1730 Lucky John Dr.
Park City, UT 84060
Phone: (319) 331-4297

Sponsor

FUJIFILM Medical Systems U.S.A., Inc.
419 West Avenue
Stamford, CT 06902

Jyh-Shyan (Jesse) Lin, PhD
Senior Manager, Regulatory, Quality and Clinical Affairs
Phone: 301.251.992
Email: jlin@fujifilm.com

Nicole Calabrese, MS
Contract Project Manager
Phone: 518-506-9840
Email: ncalabrese.contractor@fujifilm.com

Statistician

Biostatistics Consulting, LLC
10606 Wheatley Street
Kensington, MD 20895

Alicia Y. Toledano, ScD
Phone: 301.312.8577
Email: alicia.toledano@post.harvard.edu

6. Background and Rationale

Breast cancer is the most common cancer among women, excluding non-melanoma skin cancers. The chance of a woman developing invasive breast cancer at some time in her life is about 1 in 8 (12%). The American Cancer Society (ACS) estimated that in 2018 about 266,120 new cases of invasive breast cancer and about 63,960 new cases of ductal carcinoma in situ (DCIS) (the earliest form of breast cancer) would be diagnosed among women in the United States. After lung cancer, breast cancer is the leading cause of cancer death in women. The chance that breast cancer will be responsible for a woman's death is about 3%. ACS estimated that in 2018, about 40,920 women would die from breast cancer in the United States. Death rates from breast cancer have been declining since about 1989, with larger decreases in women younger than 50. These decreases are believed to be the result of earlier detection through screening and increased awareness, as well as improved treatment. [1, 2]

At present, mammography (screen-film mammography and standard full-field digital mammography [FFDM]) is the primary imaging modality used for early detection of clinically occult breast cancer. According to the 2009 report on breast cancer screening by the U.S. Preventive Services Task Force, screening mammography was associated with a reduction in the relative risk for death due to breast cancer of 14% for women aged 50 to 59 years and 32% for women aged 60 to 69 years.[3,4] In an analysis of the Swedish Mammography Screening in Young Women (SCRY) cohort, screening mammography was associated with a reduction in the relative risk for death due to breast cancer of 39% for women aged 40 to 49 years.[5]

FFDM is known to have three limitations.[6-10] These limitations can contribute to inaccurate interpretations. The first limitation is lesion masking: a true lesion may not be visible on FFDM because it is masked by superimposed overlying or underlying normal tissue. The second limitation is lesion mimicry, which can occur when normal layers of tissue or isolated elements of calcium at different depths in the breast are superimposed along the line of sight and combine on the 2D-projected image to resemble a lesion. The consequence is false-positive reports that require the unnecessary call back of subjects for further diagnostic workup examinations. The third limitation addresses the inability to perceive the in-depth structure of a detected lesion and the surrounding tissue on FFDM. This information is particularly important for suggesting the presence of spiculations and the significance of calcification distributions. For standard mammography, depth information can be obtained only in a very limited way by cognitive merging of information abstracted separately from the usual two CC- and MLO-view images of the same breast.

These limitations of FFDM may be significantly reduced through the use of digital breast tomosynthesis (DBT), which provides radiologists with in-depth views (through a series of cross-sectional views of DBT images) of the internal structure of the breast where each reconstructed image contains depth information about breast structure and composition. In the DBT examination, each view (CC and MLO) of the breast is acquired as 2D and DBT images. The 2D and DBT images are presented for the radiologist’s interpretation on a pair of FDA-cleared five-megapixel, grayscale display monitors (see Appendix A).

Synthesized view (S-View) images are used in the same way that 2D FFDM images are used as part of a 3D study, and are generally never viewed independently for screening or diagnosis. S-View images are created from a software reconstruction algorithm, allowing for reduction in radiation dose, since 2D images do not need to be acquired. S-View images can be acquired in any projection and have been shown to be capable of replacing 2D FFDM imaging in a 2D/3D combination exam. S-View images can be used when comparing to FFDM images, with the diagnosis and BI-RADS assessment made based on the combination of the S-View and tomosynthesis images. S-View images are also useful in detecting left/right asymmetry as well as in assessing the distributional aspects of calcification. If there is a finding in a synthesized image, it will be confirmed with the tomosynthesis images.

FFDM with DBT is being replaced by the synthesized 2D mammography with DBT in breast imaging to reduce radiation dose while maintaining screening outcomes. Recent clinical trials supporting the Selenia Dimensions 3D System (with C-View Module) - PMA Supplement: P080003/S001, and the SenoClaire Digital Breast Tomosynthesis System (with V-Preview) – PMA Supplement P130020/S001, demonstrated 3D digital mammography images suitable to be used in the screening and diagnosis of breast cancer as well as the reasonable assurance of safety and effectiveness of DBT plus synthesized 2D technology.[16,18]

7. Investigational Plan

7.1 Purpose

The purpose of the pivotal reader study is to evaluate the comparative accuracy of Fujifilm DBT plus S-View versus FFDM in the detection of breast cancer. The radiologists' performance metrics for the following modalities will be evaluated:

- FFDM read on the ASPIRE Bellus II workstation ("FFDM")
- Synthesized (S-View) read in conjunction with DBT on the ASPIRE Bellus II workstation ("DBT plus S-View")

This retrospective, multi-reader multi-case (MRMC) study will have an enriched sample of approximately 300 cases: 60 cancer cases and 240 non-cancer cases (including up to 48 biopsy proven benign cases and 72 recall cases). Enrichment is with respect to proportions of cancer cases, biopsy benign cases, and recall cases.

As its primary endpoint, this study is designed to evaluate whether the area under the receiver operating characteristic (ROC) curve (AUC) based on probability of malignancy (POM) scores and requiring correct lesion localization is statistically significantly non-inferior for DBT plus S-View versus FFDM. Among its secondary endpoints, the study is designed to evaluate:

1. Non-inferior and/or superior (lower) per-subject average recall rate for all non-cancer cases for DBT plus S-View versus FFDM, based on recall scores, using non-inferiority margin delta = 0.05.
2. Non-inferior per-subject average recall rate for DBT plus S-View versus FFDM for all cancer cases, based on recall score, using non-inferiority margin delta = 0.10
3. Non-inferior and/or superior per-subject average sensitivity for DBT plus S-View versus FFDM, based on BI-RADS scores requiring correct lesion localization, using non-inferiority margin delta = 0.10
4. Superior per-subject average AUC for DBT plus S-View versus FFDM, based on POM scores and requiring correct lesion localization
5. Non-inferior and/or superior per-subject average specificity for DBT plus S-View versus FFDM, based on BI-RADS scores
6. Superior (lower) per-subject average recall rate for DBT plus S-View versus FFDM for all follow-up proven non-cancer recall cases
7. Non-inferior per-lesion average sensitivity for masses, masses with calcifications, focal asymmetries, and/or architectural distortions for DBT plus S-View versus FFDM, based on BI-RADS scores requiring correct lesion localization, using non-inferiority margin delta = 0.10
8. Non-inferior and/or superior per-lesion average sensitivity for calcifications for DBT plus S-View versus FFDM, based on BI-RADS scores requiring correct lesion localization, using non-inferiority margin delta = 0.10
9. Non-inferior (margin delta = 0.05 for AUC, 0.10 for other performance metrics) and/or superior average AUC and/or other performance metric(s) for DBT plus S-View versus FFDM for subjects with dense breasts (BI-RADS breast composition categories c. The breasts are heterogeneously dense, which may obscure small

masses and d. The breasts are extremely dense, which lowers the sensitivity of mammography)

7.2 Study Endpoints

The safety and effectiveness of DBT plus S-View are both linked to the ability of radiologists interpreting DBT plus S-View images to accurately diagnose breast cancer. Each study endpoint therefore addresses both safety and effectiveness.

All of the following endpoints are measured on the FFDM read, DBT plus S-View read, and their difference. All measurements require correct lesion localization.

The **primary endpoint** is:

- Non-inferior per-subject average area under the ROC curve (AUC) for DBT plus S-View versus FFDM, based on POM scores and requiring correct lesion localization

The study will be considered to have successfully demonstrated safety and effectiveness of the Fujifilm ASPIRE Cristalle DBT plus S-View system if the per-subject average AUC for DBT plus S-View is statistically significantly non-inferior to the average AUC for FFDM at the alpha = 0.05 significance level, for non-inferiority margin delta = 0.05. This will be established if the lower limit of the two-sided 95% CI for the difference in average AUC for DBT plus S-View – FFDM lies entirely above –0.05.

The **secondary endpoints** are:

1. Non-inferior and/or superior (lower) per-subject average recall rate for all non-cancer cases for DBT plus S-View versus FFDM, based on recall scores, using non-inferiority margin delta = 0.05.
2. Non-inferior per-subject average recall rate for DBT plus S-View versus FFDM for all cancer cases, based on recall score, using non-inferiority margin delta = 0.10
3. Non-inferior and/or superior per-subject average sensitivity for DBT plus S-View versus FFDM, based on BI-RADS scores requiring correct lesion localization, using non-inferiority margin delta = 0.10
4. Superior per-subject average AUC for DBT plus S-View versus FFDM, based on POM scores and requiring correct lesion localization
5. Non-inferior and/or superior per-subject average specificity for DBT plus S-View versus FFDM, based on BI-RADS scores
6. Superior (lower) per-subject average recall rate for DBT plus S-View versus FFDM for all follow-up proven non-cancer recall cases
7. Non-inferior per-lesion average sensitivity for masses, masses with calcifications, focal asymmetries, and/or architectural distortions for DBT plus S-View versus FFDM, based on BI-RADS scores requiring correct lesion localization, using non-inferiority margin delta = 0.10
8. Non-inferior and/or superior per-lesion average sensitivity for calcifications for DBT plus S-View versus FFDM, based on BI-RADS scores requiring correct lesion localization, using non-inferiority margin delta = 0.10
9. Non-inferior (margin delta = 0.05 for AUC, 0.10 for other performance metrics) and/or superior average AUC and/or other performance metric(s) for DBT plus S-View versus FFDM for subjects with dense breasts (BI-RADS breast composition categories c. The breasts are heterogeneously dense, which may obscure small

masses and d. The breasts are extremely dense, which lowers the sensitivity of mammography)

Multiple Comparisons. To protect the study's type I error rate from inflation we will use the iterative graphical approach described in Bretz, et al. (2009) to sequentially reject hypotheses.[13]

7.3 Pivotal Study Population

A minimum of 300 cases (60 cancer cases and 240 non-cancer cases) will be selected from mammograms that were acquired during Fujifilm protocol FMSU2013-004A.

Images to be used in this study were acquired at image-acquisition sites listed in Table 1. The sites participated in protocol FMSU2013-004A and acquired images of subjects who consented to participate in that study. As part of the consent process, subjects agreed that image data and supporting documentation could be used for future research and investigations.

Table 1: Image Acquisition Sites and Principal Investigators

Image-Acquisition Site	Principal Investigator's Name	City, State
Elizabeth Wende Breast Care, LLC	Stamatia Destounis, MD	Rochester, NY
Scottsdale Medical Imaging, LLC	Denise Reddy, MD	Scottsdale, AZ
University of Iowa Hospitals and Clinics	Limin Yang, MD	Iowa City, IA
The University of North Carolina at Chapel Hill	Cherie Kuzmiak, DO	Chapel Hill, NC
University of Texas San Antonio	Pamela Otto, MD	San Antonio, TX

7.3.1 Case Reference Standard

The **reference standard for a cancer case** is a case in which at least one lesion is confirmed as malignant by biopsy or surgery.

The **reference standard for a non-cancer case** is biopsy or surgery for benign cases, or interpretation by radiologists at the acquisition site for recall and normal cases.

7.3.2 Sample Size and Case Mix

Approximately 60 cancer cases and approximately 240 non-cancer cases will be selected by the statistician for this pivotal reader study from the library of FFDM and DBT mammograms collected under protocol FMSU2013-004A. All cancer cases will undergo the truthing process described in section 8. The approximately 240 non-cancer cases will include:

- approximately 48 biopsy benign cases
- approximately 72 recall cases
- approximately 120 normal cases

7.3.3 Inclusion and Exclusion Criteria

All cases for this pivotal MRMC reader study will meet the following eligibility inclusion and exclusion criteria (see also Appendix B for inclusion and exclusion criteria from protocol FMSU2013-004A). Only subjects who were consented, met all inclusion criteria and none of the exclusion criteria, and completed the needed work-up to establish the ground truth were considered evaluable.

Inclusion Criteria

- Eligible subjects under protocol FMSU2013-004A, defined as female subjects with known true clinical status and with complete FFDM and DBT examinations, in which there is sufficient anatomical coverage, sufficient contrast, and no significant motion or other artifacts, as determined by the image-acquisition sites.
- Meet none of the exclusion criteria under protocol FMSU2013-004A.

Exclusion Criteria

- Subjects who are in violation of protocol FMSU2013-004A.
- Subjects who meet exclusion criteria under protocol FMSU2013-004A.
- Subjects with unknown clinical status.
- Any subject whose positive mammogram was not read during the truthing process (see section 8) will not be considered for the pivotal MRMC reader study.

7.3.4 Case selection and Targeted Distribution

Cases within each category of cancer, benign, recall, and normal cases will be randomly selected to provide the (targeted) representative distributions of:

- mammographic finding types (Table 2),
- breast composition (Table 3),
- and equal distribution across acquisition site, to the extent that this can be accomplished using cases in the FMSU2013-004A case library.

The statistician may provide a backup list of cases to be used in the event that, e.g., imaging files are not available for the reader study.

Table 2: Targeted distribution of mammographic finding types (approximation)

Case Type	Calcification (incl. calcification w/ Mass)	Mass (non calcification)	No Findings	Total
Cancer	10%	10%		20%
Benign	8%	8%		16%
Recall	12%	12%		24%
Normal	0	0	40%	40%
Total	30%	30%	40%	100%

Table 3: Targeted distribution of mammography breast composition (approximation)

Breast Tissue Density	BI-RADS Breast Tissue Composition Category	Total
Fatty (50%)	a - Almost Entirely Fatty	10% to 15%
	b - Scattered Fibroglandular	40% to 45%
Dense (50%)	c - Heterogeneously Dense	35% to 40%
	d - Homogeneously Dense	5% to 10%
	Total	100%

7.4 Risk/Benefit Assessment

The probable benefits of the device are based on the data that will be collected in the clinical study being conducted to support PMA supplement approval.

The ASPIRE Cristalle DBT System is used to reconstruct the breast volume from limited angle projections while eliminating the tissue overlapping effects observed in FFDM projections. It is likely to benefit a substantial number of screening patients whose cancers could have otherwise been missed due to tissue superimposition (false negatives), or who may otherwise have been unnecessarily referred for additional workup (false positive).

Synthesized view (S-View) images are created from the reconstructed DBT images. Without the need of additional radiation exposure to the patient, the S-View image is generated by “combining” the DBT images of that view into a single 2D synthetic mammography image.

The device has no significant risk of direct harm to the patient.

The primary risk of the device comes from the possibility of false positive and false negative

clinical decisions when using the images produced by the ASPIRE Cristalle DBT System. Fujifilm has designed an MRMC study to compare the performance of readers with FFDM alone and with DBT plus S-View. The study design is consistent with other mammography studies. Because MRMC studies are conducted outside the clinical setting, with an enriched case set, and without patient history, the generalizability of some figures of merit such as recall rate, sensitivity, and specificity is limited. The design is considered acceptable in order to reduce the size of the trial and avoid confounders.

8. The Truthing Process

8.1 Purpose

The purpose of the truthing process will be to determine, for the cancer cases selected by the statistician for this pivotal reader study, the type of lesion (or finding) visible within the mammogram, the location for each lesion (or finding), and the pathology results.

8.2 Qualification Criteria and Training for the Truther

The truther is a radiologist who meets the expert radiologist eligibility criteria (below), is not associated with the diagnosing of cases at the image-acquisition sites (under Fujifilm protocol FMSU2013-004A), will not be associated with the reading of images in this pivotal reader study, and will be available to perform the responsibilities as the study truther and conduct the truthing process.

Base qualification criteria: The base criteria for the truther are as follows:

- Board-certified radiologist.
- MQSA-qualified in tomosynthesis
- Experience reading FFDM images on a monitor (softcopy).
- Experience reading DBT images on a monitor (softcopy).

In addition, to merit the designation “expert in breast imaging,” the radiologist selected to be the study truther must meet one or more of the following criteria:

- Have instructed breast imaging in a professional society meeting.
- Have authored textbooks, book chapters, or peer-reviewed journal articles in mammography.
- Have interpreted mammography for at least 10 years.

Training: The truther will participate in a training session to ensure understanding of the ASPIRE Bellus II workstation, reading DBT plus S-View images, the protocol, the eCRFs, study processes, GCP, and the obligations entailed by study participation. Particular attention will be given to reviewing the annotation process.

The study truther will be available as necessary to assist in the conduct of the pivotal reader study. Possible services to be provided by the study truther include the following:

- Conduct of the truthing process (see section 8.3) for this clinical study.
- Review of some or all cases selected to be read during a pivotal reader study.
- Assistance with any clinical questions pertaining to the clinical investigation.
- Designation of the location and view(s) in which a finding or findings is/are located. The type of lesion (finding) visible within the mammogram and location(s) designated and annotated by the truther will be considered as establishing the true type and location of the findings.

8.3 Conduct of the Truthing Process

8.3.1 Core Reading Center

Truthing will take place at International HealthCare, LLC in Norwalk, CT.

8.3.2 Preparation of Cases for Truthing

The study truther may have access to basic information about the subjects including age at exam, brief pertinent medical history, BI-RADS descriptors, finding location, and final clinical assessment. The anonymity of participating subjects will be maintained. See section 12.6.1 for detail about maintenance of subject confidentiality, study ID numbers, and tracking.

Truthing will be performed in a controlled environment with consistent lighting. Using the ASPIRE Bellus II workstation, the truther will respectively review the FFDM and DBT plus S-View images. During the review, the truther will be permitted to use standard softcopy imaging tools and all routine post-processing capabilities (for example, magnification and windowing/leveling).

8.3.3 Workflow for the Truthing Process

The truther will review all cancer cases selected for inclusion in the reader study. The truther will complete an eCRF for each case, based on materials from the image acquisition site including mammograms. For each cancer case the truther will review the images and identify, confirm, and annotate (including the lesion locations and slice, if applicable) for each malignant lesion:

- The type of lesion visible within the mammogram (mass, microcalcifications asymmetric density and/or focal asymmetry)
- The breast, view, and if possible the coordinates and slice, if applicable, in which each lesion is located
- Document the characteristics of the lesion on the eCRF

9. Reader Selection and Training

9.1 Purpose

The purpose of the training process is to ensure that each reader selected for the reading session has been adequately trained to read and interpret the DBT plus S-View images on the ASPIRE Bellus II workstation. A radiologist with expertise in digital breast tomosynthesis will conduct the training of the radiologists.

9.2 Selection and Qualification Criteria for Readers

This study will be conducted with radiologists of varying experience levels, from both community and academic practices; some of whom will have had no experience reading synthesized (S-View) images. These radiologists will not be associated with the acquisition or review of cases and did not participate in the truthing process. A survey questionnaire will be sent to prospective readers to determine the fraction of their professional time devoted to breast imaging. Approximately 18 board certified radiologists, MQSA qualified in tomosynthesis and FFDM and with a range of experience based on the volume of cases read per year, will be selected to read and score the study images.

Case volume will be categorized into three groups:

- High Volume Reader - defined as radiologists who read more than 5000 mammograms per year.
- Medium Volume Reader - defined as radiologists who read more than 3000 up to 5000 mammograms per year.
- Low Volume Reader - defined as radiologists who read 3000 or fewer mammograms per year.

Base qualification criteria: All readers must meet all of the following base criteria:

- Are board-certified
- Are MQSA-qualified for both FFDM and DBT interpretation

9.3 Conduct of the Training Process

The radiologists will receive approximately three hours of training in the evaluation of DBT plus S-View images to familiarize them with the S-View images. Training will consist of the trainer presenting a brief didactic lecture explaining how synthetic 2D images are generated and the differences, advantages and pitfalls compared with digital mammography. The trainer will demonstrate how to use and interpret S-View images by presenting selected cases to demonstrate the range of typical anatomy, breast density and abnormalities identified with DBT. Training will also consist of a hands-on session at the workstation to provide the readers with an overview of its DBT-specific functionality. For each view, mediolateral oblique (MLO) and cranio-caudal (CC), 2D FFDM and the corresponding S-View image for the same view will be shown with the DBT images. The training will also emphasize that the S-View images alone will not be used for diagnosis, and scoring will be based on the appearance of the lesion on the DBT images.

10. The Reading Process

10.1 Purpose

The purpose of the reading process is for the readers to provide the BI-RADS assessment categories, probability of malignancy (POM) scores, recall scores, locations and types of any finding(s) on the eCRFs for their interpretation of the study cases; with FFDM images and DBT plus S-View images on the ASPIRE Bellus II workstation.

10.2 Conduct of the Reading Process

Following the successful completion of the reader training, the readers will independently review and score each of the image sets of the reader study case set. Each reader will work at a designated workstation and will independently perform two reads on all cases. Each reader will read each case both as a FFDM read, and a DBT plus S-View read on the ASPIRE Bellus II workstation. Each reader will read half the cases on FFDM and the other half of the cases on DBT **p**lus S-View during the first visit. The reader will then read the complementary FFDM and DBT plus S-View during the second visit. In order to minimize memory recall bias, the two reading visits will be separated by a memory washout period of approximately four weeks.

Interpretation of mammography images will be performed in a controlled environment with consistent lighting. The review workstations will be preloaded with the cases. The readers will be permitted to use standard softcopy imaging tools, and all routine post processing capabilities (e.g., magnification, windowing and leveling) will be permitted. Hotkeys will be provided on the workstations for consistency and accuracy of the reading process. The reading time allowed per case will not be limited. A lunch break will be offered during each reading session, and radiologists will be encouraged to take a 5-10 minute break each hour, as needed. Each reader will dictate their interpretations to a trained scribe, who will enter the dictation directly into the eCRF. Each reader will sign an attestation form verifying the documented information.

10.2.1 Core Reading Center

The reading process will be conducted at International HealthCare, LLC in Norwalk, CT.

10.2.2 Preparation of Cases for Interpretation

To maintain subject confidentiality, no identifying subject data will be displayed. Unique image identifiers associated with the case ID number will be visible on each image for tracking purposes. To maintain reader confidentiality, a unique ID number will be assigned to each radiologist participating in the pivotal reader study for entry on the CRFs. Information collected during the research study was reported in such a way that it precluded identification of any image acquisition centers' or study radiologists' performance.

10.2.3 Workflow for the Interpretation of Cases

The readers will be told that the samples of cases do not represent a standard screening population, and will be blinded to the actual distribution and nature of the set of images they will be asked to review. Readers will be masked to the reference standard and image acquisition interpretations (under protocol FMSU2013-004A) for each case. Readers will not have access to prior mammograms or other clinical information. All readers will perform their interpretations independently.

Each study reader will be randomly assigned to a unique reading order. All study readers will ultimately interpret all study cases in both FFDM, and DBT plus S-View formats. Readers will interpret the preloaded mammograms in their assigned order with the use of barcode scanner/barcodes.

General instructions for the readers are as follows:

- Do not record single punctuate or round calcifications and vascular calcifications, given the high prevalence of these benign findings and their low probability of cancer.
- Consider all views available from both breasts to determine if you would recall the case.
- For examinations with multiple findings, record the types and locations of all findings in the eCRF and/or a breast schematic diagram that divides each breast image into sections and image slice for DBT images. Record the finding locations beginning with the most suspicious lesion (up to three lesions per case).
- Provide a BI-RADS assessment category of 1 through 5 for each annotated finding.
- Provide a probability of malignancy (POM) for each annotated finding according to the guidance listed in Table 4 (below).

Probability of Malignancy (POM): POM is a 0 to 100 score assigned by readers to each suspicious finding, as their perception of the percentage chance that the suspicious finding might be malignant. To help the assignment of POM scores readers will be instructed to use the entire POM scale, with descriptions of POM values provided in Table 4 only as guidance:

Note: POM scores will be determined and analyzed separately from the BI-RADS assessment category.

Table 4: Probability of malignancy (POM) guidance

POM Score (%)	Description
0-20	Negative/Benign
21–40	Probably benign
41–60	Possibly malignant
61–80	Probably malignant
81–100	Malignant

Interpretation instructions for the readers are as follows:

The reader will complete the eCRFs for each case for the two reads. For each case on each read, the reader will first record whether there are mammographic findings. If the answer to this question is “no” the reader will be asked to confirm for this case: a BI-RADS assessment category of 1 or 2, a POM score following the POM guidance provided in Table 4, and recall decision of “no”. If the reader answers “yes” to whether there are mammographic findings, the reader will be asked to confirm an initial BI-RADS assessment category of 0, and will then provide detailed information on each suspicious finding:

- Affected breast (right or left)
- Type (includes mass, asymmetry, and/or architectural distortion), microcalcifications, or mass with microcalcifications, or other
- “Forced” BI-RADS assessment category 1, 2, 3, 4, or 5
- POM score 0 through 100
- The reader will then be asked whether the case shall be recalled (“yes” or “no”), their forced BI-RADS assessment category (1, 2, 3, 4 or 5), and POM score 0 through 100.

A study scribe representing the sponsor will capture the radiologist’s results on the eCRFs to ensure completeness.

In cases with mammographic findings, consistency of POM scores with BI-RADS assessment categories and recall decisions will not be forced – i.e. readers will be permitted to use the full range of POM scores for a finding (following the POM guidance provided in Table 4) no matter what BI-RADS assessment category they assign to it.

10.2.4 Reading Sessions with a Memory Washout Period

Each reader will read each case both as a FFDM read and a DBT plus S-View read, on the ASPIRE Bellus II workstation. Each reader will read half the cases on FFDM and the other half of the cases on DBT plus S-View during one visit. The reader will then read the complementary FFDM and DBT plus S-View during the second visit. In order to minimize memory recall bias, the two reading visits will be separated by a memory washout period of approximately four weeks.

Reading Session Order

a. The selected cases will be randomly allocated into four (4) sets of 75 cases each, case subsets A, B, C, and D. Allocation will be balanced with respect to reference standard (cancer, benign, recall, and normal), and to the extent possible in the case sample, on presence of calcifications, breast composition (fatty or dense), and image acquisition site.
b. Each study reader will be randomly assigned to a unique reading order.
c. Each session has 300 cases that will be read: **150** FFDM cases and **150** DBT plus S-View cases for the two read visits.

Table 5: One Possible Experimental Design

	Session 1					Session 2				
Pat.	Part 1		Part 2			Part 1		Part 2		Reader
	Gr. 1	Gr. 2	Gr. 1	Gr. 2		Gr. 1	Gr. 2	Gr. 1	Gr. 2	
1	B FFDM	D DBT +	C DBT +	A FFDM	W	A DBT +	C FFDM	D FFDM	B DBT +	11, 17
2	A FFDM	C DBT +	D DBT +	B FFDM	A	B DBT +	D FFDM	C FFDM	A DBT +	10, 20
3	C FFDM	B DBT +	A DBT +	D FFDM	S	D DBT +	A FFDM	B FFDM	C DBT +	14, 21, 24
4	D FFDM	A DBT +	B DBT +	C FFDM	H	C DBT +	B FFDM	A FFDM	D DBT +	9, 19
5	B DBT +	D FFDM	C FFDM	A DBT +	O	A FFDM	C DBT +	D DBT +	B FFDM	8, 15
6	A DBT +	C FFDM	D FFDM	B DBT +	U	B FFDM	D DBT +	C DBT +	A FFDM	7, 16
7	C DBT +	B FFDM	A FFDM	D DBT +	T	D FFDM	A DBT +	B DBT +	C FFDM	12, 18
8	D DBT +	A FFDM	B FFDM	C DBT +		C FFDM	B DBT +	A DBT +	D FFDM	13, 22, 23

Pat. = Pattern. Gr. = Group. Reader numbers begin with 7 to prevent confusion with reader numbers on the pilot study. DBT + = DBT plus S-View.

Interpretation Results: For each case, the readers are asked to provide two sets of results from two reading visits separated by a memory washout period. Each reader will read half the cases on FFDM and the other half of the cases on DBT plus S-View during the first visit. The reader will then read the complementary FFDM and DBT plus S-View during the second visit.

11. The Lesion Matching Process

11.1 Purpose

The purpose of the lesion matching process will be to determine if the location and type of a finding identified by the reader matches those of a malignant lesion annotated by the truther, i.e., correct lesion localization and type.

11.2 Qualification Criteria and Training for the Lesion Matcher

An expert not associated with diagnosing cases at the image-acquisition sites or serving as a reader interpreting images on this pivotal reader study will serve as the lesion matcher for this study. This expert may be an experienced, MQSA-qualified mammography technologist with an active license, or a board-certified radiologist that is MQSA-qualified for both FFDM and tomosynthesis; if the latter, this expert radiologist may be the same person as the truther.

Base qualification criteria: The base criteria for the lesion matcher are as follows:

- Has an active license as a breast mammogram technologist
- Has experience with clinical research in the GCP environment
- Or, Study truther (refer to truther qualifications in section 8.2)

Training: The lesion matcher will participate in a training session to ensure understanding of the ASPIRE Bellus II workstation, the protocol, the eCRFs, study processes, GCP, and the obligations entailed by study participation. Particular attention will be given to the lesion matching process.

11.3 Conduct of the Lesion Matching Process

11.3.1 Core Matching Center

Lesion matching will take place remotely, with results documented within the truther/lesion matching database.

11.3.2 Preparation of Cases for Lesion Matching

The lesion matcher will have access to both the reader's eCRF and the truther's eCRF. The eCRF and images contain the annotated information, such as Subject ID number, the finding location(s), and finding type(s) which are required for the determination of correct lesion localization. The matching is limited to cases with malignant lesions. For each case with cancer, the lesion matcher determines if the annotation by the reader matches with that made by the truther.

11.3.3 Workflow for the Lesion Matching Process

The lesion matching process is limited to those cases with malignant lesions. Based on the case ID number on the truther's eCRF, the lesion matcher identifies the reader's eCRF with the same Subject ID number. The lesion matcher compares the finding location(s) and finding type(s) of the reader's eCRF and annotated images to those of the truther's eCRF, to

determine if the reader made the correct lesion localization. The process will be repeated for each reader's eCRFs that have the corresponding Reader Sequence ID number to truther's eCRFs, and for every reader. The lesion matcher will record the lesion matching results on the eCRFs.

- For each truther's eCRF, identify and confirm the reader's eCRF that contains the corresponding Subject ID number and Reader Sequence ID Number
- Identify and match the location (breast, view, and if possible the section and/or slice) and type of finding(s) with those of the truther's eCRF
- Document the matching results on the eCRF

Note 1: To allow for the possibility of lesions located on breast section boundaries, the reader will get credit for correctly locating a lesion if the matcher determines that the annotations match even if an adjacent quadrant is recorded on the case report form. In this situation, we anticipate x/y coordinates to be within 2 centimeters and slice numbers to be within 10 slices. Readers will not get credit for correctly locating a lesion if the matcher determines that the annotations do not match, even if the same quadrant is recorded on the case report form; in this situation the x/y coordinates and/or slice numbers also usually would not match with the pre-specified tolerance. The lesion matching process will be supported by programmatic (computer) determination of whether x and y coordinates are within 2 centimeters; and whether slice numbers are within 10 slices.

12. Statistical Analysis

Study case identification numbers and study reader numbers will be assigned to all cases and readers, and used to protect their identities in statistical analysis and in reporting of results. All cases will be evaluated the same way by all study readers, that is, there will be no treatment assignments or treatment groups. No interim analyses are planned. No adverse events are anticipated on this MRMC study using retrospective cases. Test reproducibility will not be evaluated on this MRMC study. The statistician will not be blinded to the review setup, because slice numbers apply only to DBT plus S-View.

Baseline descriptive summaries will include the distribution of demographic characteristics and type of cancer (soft tissue lesion [mass, asymmetry, architectural distortion, other], microcalcification, or soft tissue lesion with microcalcification), lesion size, breast composition (BI-RADS breast composition categories) and reference standard status (cancer, benign, recall, or normal). We will also provide summaries across readers of the per-subject number of findings, BI-RADS scores, POM scores, and per-subject recall scores, for the FFDM and DBT plus S-View workstation image reviews. These may be cross-classified by, for example, presence of malignant lesions. Categorical variables (such as cancer type and breast tissue composition) will generally be summarized using frequencies and proportions or percentages, while continuous variables will generally be summarized using means and standard deviations (SDs), or medians and quartiles or ranges. Missing values will generally be reported as such in these descriptive summaries.

12.1 Analysis Set

Intent-to-Diagnose Population The intent-to-diagnose population comprises each reader's interpretation of each study case in each modality (FFDM, DBT plus S-View). Missing interpretations are not anticipated. We plan to include all readers' interpretations of all cases in both modalities in the analysis set. If any protocol deviations or violations occur, the statistician will evaluate them to determine their impact on the integrity of the study database; and will determine whether any affected data points should be excluded from primary analysis.

12.2 Multiple Centers (pooling)

Fujifilm has obtained images from multiple centers. The protocol for data submission, quality review passed by all images, and reference standard status determination for all images used in the MRMC study will be common. Cases will be pooled across enrolling centers for interpretation on the MRMC study using common interpretation protocol and electronic case report forms (eCRFs), and results of interpretation sessions will be monitored. The scoring of (lesion matching for) reader interpretations will follow a common process. Results for any particular reader will therefore be pooled across enrolling centers.

12.3 Derived Variables for Statistical Analyses

The per-subject BI-RADS score, POM, and recall scores requiring correct lesion localization will be derived as in sections 12.3.1 and 12.3.2, respectively. The general principle is that even at the subject level, credit is only given for identifying a subject with cancer if the reader marks findings in at least one location with cancer. Findings that do not match the location of a malignant lesion are ignored for cancer cases in the per-subject analyses but may be reported, generally in an appendix.

12.3.1 Per-subject scoring: POM and BI-RADS

The primary endpoint is per-subject AUC based on POM scores requiring correct lesion localization. Secondary endpoints include per-subject sensitivity and specificity based on BI-RADS categories. When computing sensitivity and specificity based on BI-RADS, a score of 4 or 5 constitutes a positive test result. A cutoff score of BI-RADS 3 or higher may also be used to compute the sensitivity and specificity in secondary analyses. Scores for use in these per-subject analyses will be derived by the statistician as described below and summarized in Table 6:

- If a case has no proven malignancies:
    - If the reader does not record any findings in this case: The POM score for this case will be the same as the POM score recorded by the reader for the case, and the BI-RADS score for this case will be the same as the BI-RADS assessment category recorded by the reader for the case.
    - If the reader records any findings in this case: The overall POM score recorded by the reader for the case, and the overall BI-RADS category recorded by the reader for the case will be used, after describing those findings.
- If a case has one or more proven malignancies:
    - If the reader records one or more findings that are determined by the truther to match the location(s) of **any** proven malignancies in this case: The maximum of the POM scores recorded by the reader for those findings will be assigned to the case, and the BI-RADS category for the case will be assigned as the maximum of the BI-RADS categories recorded by the reader for those findings.
        - If the case contains more than one malignant lesion, the reader will get credit for identifying the case as having one or more proven malignancies even if the reader does not identify all of the proven malignancies in the case. For example in a bilateral case, the reader would get credit for identifying the case even if the reader marks findings in only one breast.
    - If the reader does not record any findings in this case; or when the reader records findings in this case, but none of them are determined by the lesion matcher to match the location(s) of any proven malignancies:
        - If the reader did not record any findings in the case, the POM score for this case will be the same as the POM score recorded by the reader for the case and the BI-RADS score for this case will be the same as the BI-RADS assessment category recorded by the reader for the case
        - If the reader recorded one or more findings in this case, but none of them are determined by the lesion matcher to match the location(s) of any proven malignancies, **a POM score of the higher of 0 or, for readers who do not assign POM 0 to any case in a reading modality, the minimum POM score assigned by that reader in**

that modality will be assigned to the case and the BI-RADS score for this case will be assigned as category 1.

- The POM scores and BI-RADS categories for any reader findings in this case that do not match the location(s) of any proven malignancies will be ignored in the per-subject analysis, which requires a single POM score and single BI-RADS category per subject conditional on whether the subject does or does not have proven malignancies.

Table 6: Per-subject POM and BI-RADS scores (requiring correct lesion localization)

Reference standard	Reader's interpretation	Per-subject POM and BI-RADS Scores
No malignancies in this case	No findings located in this case	**POM**: Same as POM recorded by the reader for the case **BI-RADS**: Same as BI-RADS category recorded by the reader for the case
	One or more findings located in this case	**POM**: Overall POM score recorded by the reader for the case **BI-RADS**: Overall BI-RADS category recorded by the reader for the case
One or more malignancies in this case	No findings located in this case	**POM**: Same as POM recorded by the reader for the case **BI-RADS**: Same as BI-RADS category recorded by the reader for the case
	Findings in this case, but no findings matching the location(s) of any proven malignancies in this case	**POM**: Assigned as the higher of zero (0) or, for readers who do not assign POM 0 to any case in a reading modality, the minimum POM score assigned by that reader in that modality. **BI-RADS**: Assigned as category one (1)
	One or more findings correctly matching the location(s) of any proven malignancies in this case	**POM**: Highest POM score recorded by the reader for any of those matching findings **BI-RADS**: Highest BI-RADS category recorded by the reader for any of those matching findings

12.3.2 Per-subject scoring: Recall

The secondary endpoints include per-subject recall requiring correct lesion localization based on a separate yes/no question. Scores for use in these per-subject analyses will be derived by the statistician as described below and summarized in Table 7:

- If a case has no proven malignancies:
  - If the reader does not initially recall the case; or if the reader initially recalls the case and then records overall recall decision no: The recall score for this case will be the same as the recall score recorded by the reader for the case, that is, no recall.
  - If the reader initially recalls the case and records overall recall decision yes: The recall score for this case will be the same as the recall score recorded by the reader for the case, that is, recall.

- If a case has one or more proven malignancies:
  - If the reader initially recalls the case, records one or more findings that are determined by the truther to match the location(s) of any proven malignancies in this case, and records overall recall decision yes: The recall score for this case will be the same as the recall score recorded by the reader for the case, that is, recall.
  - If the reader does not initially recall the case; or if the reader initially recalls the case and then records overall recall decision no; or when the reader initially recalls the case and records overall recall decision yes, but none of the reader's findings are determined by the truther to match the location(s) of any proven malignancies:
    - If the reader did not recall the case, the recall score for this case will be the same as the recall score recorded by the reader for the case, that is, no recall.
    - **If the reader recalled the case and recorded one or more findings in the case, but none of them are determined by the truther to match the location(s) of any proven malignancies, the recall score for this case will be assigned as no recall.**

Table 7: Per-subject recall scores (requiring correct lesion localization)

Reference standard (truth) for this case	Reader's interpretation	Per-subject recall score
No malignancies	No recall (initially or overall)	Same as recall recorded by the reader for the case, that is, no recall
	Recall (initially and overall)	Same as recall recorded by the reader for the case, that is, recall
One or more malignancies	No recall (initially or overall)	Same as recall recorded by the reader for the case, that is, no recall
	Recall (initially and overall); Findings in this case, but no findings matching the location(s) of any proven malignancies in this case	**Assigned as no recall**
	Recall (initially and overall); One or more findings correctly matching the location(s) of any proven malignancies in this case	Same as recall recorded by the reader for the case, that is, recall

12.3.3 True Positive, False Negative, True Negative, and False Positive

In **per-subject analysis**:

- A true positive (TP) occurs when a case contains one or more cancerous lesions, and the per-subject BI-RADS score is 4 or 5. This happens when the reader marks one or more findings, at least one of those findings matches the location of one or more cancerous lesions, and the reader assigns a BI-RADS category 4 or 5 to at least one such finding.

- A false negative (FN) occurs when a case contains one or more cancerous lesions, and the per-subject BI-RADS score is 1, 2, or 3. This happens when either: 1) the reader does not mark any findings that match the location of one or more cancerous lesions, or 2) the reader marks one or more findings, at least one of those findings matches the location of one or more cancerous lesions, but none of the scores assigned to those findings are higher than BI-RADS 3.
- A true negative (TN) occurs when a case does not have any cancerous lesions, and the per-subject BI-RADS score is 1, 2, or 3. This happens when either: 1) the reader does not mark any findings in the case, or 2) the reader marks one or more findings in the case, but none are higher than BI-RADS 3.
- A false positive (FP) occurs when a case does not have any cancerous lesions, and the per-subject BI-RADS score is 4 or 5. This happens when the reader marks findings in a case containing no cancerous lesions, and assigns a BI-RADS category 4 or 5 to at least one such finding.

When computing recall rates requiring correct lesion localization, a recall occurs when a case has per-subject recall score equal to yes. This happens whenever the reader recalls a non-cancer case; or when the reader recalls a cancer case and marks one or more findings and at least one of those findings matches the location of one or more cancerous lesions.

12.4 Subgroups

Analysis of per-subject and per-lesion sensitivity is limited to the subgroup of subjects with cancer. Sensitivity will be analyzed in subjects with cancer overall, and may be analyzed in further subgroups defined by lesion types (for example, masses, calcifications). Recall rate also will be analyzed in the subgroup of subjects with cancer.

Analysis of per-subject specificity is limited to the subgroups of subjects without cancer. Recall rate also will be analyzed in the subgroups of subjects without cancer, and may be analyzed in the subset of recall cases. If analysis is performed in the subset of recall cases, complementary analysis will be performed in the subset of normal cases and in the subset of benign cases.

Performance metrics (AUC, sensitivity, specificity, and/or recall rate) also may be analyzed in the subgroup of women with dense breasts (BI-RADS breast composition categories c. The breasts are heterogeneously dense, which may obscure small masses and d. The breasts are extremely dense, which lowers the sensitivity of mammography). If analysis is performed in the subset of women with dense breasts, complementary analysis will be performed in the subset of women with non-dense breasts.

12.5 Analysis of Primary Endpoint

The primary endpoint on this study is per-subject AUC requiring correct lesion localization. Primary analysis will not involve pooling across study radiologists, to allow for heterogeneity among them. We will estimate AUCs for each reader in each review condition (FFDM, DBT plus S-View) based on per-subject POM scores as described in section 12.3. These POM scores will require correct lesion localization, such that in a subject with cancer if the reader recorded one or more findings in the case but no findings are determined by the lesion matcher to match the location(s) of any proven malignancies, a POM score of 0 will be assigned to the breast.

We will provide graphs of each reader's ROC curve for each review condition, and an average empirical ROC plot within each review condition. For each reader the non-parametric (trapezoidal) AUC for the FFDM read, the DBT plus S-View read, and the

difference between them, will be presented. Statistical inferences will account for correlations arising from having all study readers interpret all study cases. We plan to perform MRMC comparison of AUCs between FFDM and DBT plus S-View using the standard MRMC analysis of variance (ANOVA) method of Obuchowski and Rockette (1995)[11] to ensure generalization of the study results both to the population of readers and the population of cases. Two-sided 95% confidence intervals (CIs) will be used to quantify uncertainty in the within-modality estimates and the between-modalities difference.

The study will be considered to have successfully demonstrated safety and effectiveness of the Fujifilm ASPIRE Cristalle DBT plus S-View system if the per-subject average AUC for DBT plus S-View is statistically significantly non-inferior to the average AUC for FFDM at the alpha = 0.05 significance level, for non-inferiority margin delta = 0.05. This will be established if the lower limit of the two-sided 95% CI for the difference in average AUC for DBT plus S-View – FFDM lies entirely above –0.05.

12.6 Analysis of Secondary Endpoints

When analyzing secondary aims per-subject POM, BI-RADS, and recall scores requiring correct lesion localization will be derived as described in 12.3.1 and 12.3.2. Two-sided 95% CIs will be used to quantify uncertainty.

1. Non-inferior and/or superior (lower) per-subject average recall rate for all non-cancer cases for DBT plus S-View versus FFDM, based on recall scores, using non-inferiority margin delta = 0.05.

   Analysis of per-subject recall rates for all non-cancer cases will be performed using standard MRMC analysis of variance (ANOVA) methods described in Obuchowski and Rockette (1995).[11]
2. Non-inferior per-subject average recall rate for DBT plus S-View versus FFDM for all cancer cases, based on recall score, using non-inferiority margin delta = 0.10

   Analysis of per-subject recall rate for all cancer cases also will use the MRMC method of Obuchowski and Rockette (1995).[11]
3. Non-inferior and/or superior per-subject average sensitivity for DBT plus S-View versus FFDM, based on BI-RADS scores requiring correct lesion localization, using non-inferiority margin delta = 0.10

   Analysis of per-subject sensitivity is limited to the subgroup of subjects with cancer. This analysis will be performed similarly to analysis of per-subject recall rates for all cancer cases.
4. Superior per-subject average AUC for DBT plus S-View versus FFDM, based on POM scores and requiring correct lesion localization

   Analysis will be performed similarly to analysis of the Primary Aim. Superiority will be established if the lower limit of the two-sided 95% CI for the difference in average AUC for DBT plus S-View – FFDM lies entirely above zero (0).
5. Non-inferior and/or superior per-subject average specificity for DBT plus S-View versus FFDM, based on BI-RADS scores

   Analysis of per-subject specificity is limited to the subgroup of subjects without cancer. This analysis also will be performed similarly to analysis of per-subject sensitivity.

6. Superior (lower) per-subject average recall rate for DBT plus S-View versus FFDM for all recall cases (follow-up proven non-cancer cases)

   Analysis of per-subject recall rate for all recall cases will be performed similarly to analysis of per-subject recall rates for all non-cancer cases. If analysis is performed in the subset of recall cases, complementary analysis will be performed in the subset of normal cases and in the subset of benign cases.

7. Non-inferior per-lesion average sensitivity for masses, masses with calcifications, focal asymmetries, and/or architectural distortions for DBT plus S-View versus FFDM, based on BI-RADS scores requiring correct lesion localization, using non-inferiority margin delta = 0.10

   We plan to use methods for clustered data from MRMC studies that take into account the correlation between lesions in the same case when analyzing lesion-level sensitivity. In particular, Rao and Scott's (1992)[14] method for estimating proportions from clustered data will be used to obtain estimates for each reader in each reading condition, and Obuchowski's (1998)[12] extension of this to a pair of correlated proportions will be used to estimate the variance-covariance matrix of all possible pairs of proportions. The usual Obuchowski and Rockette (1995)[11] MRMC method will then be applied to perform inferences that generalize to the population of readers and the population of cases while also taking into account within-case correlations between lesions.

8. Non-inferior and/or superior per-lesion average sensitivity for calcifications and/or masses with calcifications for DBT plus S-View versus FFDM, based on BI-RADS scores requiring correct lesion localization, using non-inferiority margin delta = 0.10

   Analysis of per-lesion sensitivity for calcifications and/or masses with calcifications will be performed similarly to analysis of per-lesion sensitivity for masses, masses with calcifications, focal asymmetries, and/or architectural distortions.

9. Non-inferior (margin delta = 0.05 for AUC, 0.10 for other performance metrics) and/or superior average AUC and/or other performance metric(s) for DBT plus S-View versus FFDM for subjects with dense breasts (BI-RADS breast composition categories c. The breasts are heterogeneously dense, which may obscure small masses and d. The breasts are extremely dense, which lowers the sensitivity of mammography).

   Analysis will performed similarly to corresponding analysis above. If analysis is performed in the subset of women with dense breasts, complementary analysis will be performed in the subset of women with non-dense breasts.

Multiple Comparisons. We will use a graphical approach to illustrate relationships among endpoints and protect the study's type 1 error rate from inflation.[13,17] The testing strategy is shown in a figure with vertices (nodes) for each hypothesis to be tested and directed paths (arrows) between vertices. All study hypotheses to be formally tested as potential marketing claims are shown in the graph. Hypotheses are denoted H_1, H_2, …, H_k, …, H_m, where m is the total number of hypotheses to be tested. Each hypothesis to be tested is allocated initial endpoint-specific alpha, α_k, and the sum $\alpha_1 + \alpha_2 + \ldots + \alpha_k + \ldots + \alpha_m = \alpha$ for the study overall, 0.05. Initial endpoint-specific alpha can be 0 for endpoints of lesser importance; hypotheses for these endpoints will receive alpha from hypotheses that are rejected and direct alpha toward them.

Each directed path is assigned a weight between 0 and 1 indicating how much of the endpoint-specific alpha moves along the path when a hypothesis is rejected. The sum of weights leaving any hypothesis equal to 1, that is, all of the preserved alpha is used in receiving hypotheses. Paths may include a loop-back feature whereby if only one of a pair of looping hypotheses is rejected at its endpoint-specific α_k, that α_k loops back to the other hypothesis to increase its endpoint-specific alpha. Testing in a strategy with loop-back can start with any vertex that has initial endpoint-specific $\alpha_k > 0$, and all such vertices can be tested until one is found for which the null hypothesis is rejected; then testing follows the arrows. Finally, conditional passing of alpha is shown by paths with negligible weights epsilon (ε); this places higher priority on other hypotheses until those have been tested.

The graph is continually updated each time a null hypothesis is successfully rejected, as follows:

1. Pass α_k from successful H_k according to the path weights.
2. Eliminate the vertex for H_k.
3. Connect all incoming arrows to outgoing arrow tails of the deleted vertex.
4. Adjust path weights based on relative weights of previous parts of path. Maintain: a) Sum of endpoint-specific $\alpha_k = \alpha$ and b) Sum of outgoing weights from each vertex = 1.
5. If a new path duplicates an existing path, combine them and add their weights.

On this study we are interested in testing the following hypotheses:

Table 8: Hypotheses to be tested

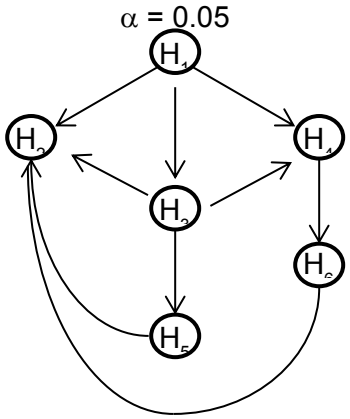
Hypothesis	Endpoint	Null Hypothesis	Alternative Hypothesis	Non-inferiority margin, delta
H_1	AUC	Inferior	Non-inferior	0.05
H_2	Per-subject recall rate for all non-cancer cases	Inferior	Non-inferior	0.05
H_3	Per-subject recall rate for all cancer cases	Inferior	Non-inferior	0.10
H_4	Per-subject sensitivity	Inferior	Non-inferior	0.10
H_5	Per-subject recall rate for all non-cancer cases	Equal	Superior	N/A
H_6	Per-subject specificity	Inferior	Non-inferior	0.05
H_7	AUC	Equal	Superior	N/A
H_8	Per-subject sensitivity	Equal	Superior	N/A

- H_1 corresponds to the study's primary endpoint. Testing will use the full study alpha: $\alpha_1 = \alpha = 0.05$. Hypothesis testing for secondary endpoints will only proceed if H_1 is rejected, in which situation α_1 will be passed to one or more of H_2, H_3, and/or H_4 through path weights w_{12}, w_{13}, and w_{14}, respectively. These weights will sum to 1, and one or two of them may be 0.
- H_2, H_3, and H_4 correspond to higher priority secondary endpoints. Their initial endpoint-specific alphas are zero, because they are only tested if the study's primary aim is met. If

that occurs their endpoint-specific alphas are updated to $\alpha_k = w_{1k} \times \alpha_1$. Each of H_2, H_3, and H_4 may pass alpha to either of the others, and this may involve loop-back. One or more of H_2, H_3, and/or H_4 also may pass alpha to one or more of H_5 through H_8, and this may be conditional on first testing all of H_2, H_3, and H_4.

- H_5 through H_8 correspond to secondary endpoints with lower priority and/or likelihood of success. Their initial endpoint-specific alphas are zero. Each of these may receive alpha from H_2, H_3, and/or H_4. Each of H_5 through H_8 may pass alpha to the others, and this may involve loop-back.
- All path weights will be provided in a separate, pre-specified analysis plan, with weights finalized prior to analyzing any study reading data. Paths with weight 0 will not appear in the final graphic.

Figure A. One example of a graphical method to control study type I error rate.



12.7 Sample Size Calculations

A sample size of at least **60** cancer cases, **240** non-cancer cases, and **18** readers was selected for this study. Sample sizes were calculated to show non-inferior AUC on a per-subject basis (primary endpoint).

We used the method of Obuchowski (1995, 2000)[12,19] to determine the number of readers required in a fully crossed design to provide 80% power at statistical significance level alpha = 0.05 for the multi-reader, multi-case (MRMC) F*-test when the number of positive cases (cancers) is **60** and the number of non-cancer cases is **240**. Calculations were made based on the following assumptions:

- Endpoint: Difference in average area under ROC curve (A) for DBT plus S-View versus FFDM
- Significance level alpha (α) = 0.05

- Target power = 0.80
- Non-inferiority margin delta (δ) = 0.05
- Average area under ROC curve (A) for FFDM, A_1 = 0.80
- Under the null hypothesis of inferiority, the average area under the ROC curve for DBT plus S-View is A_2 = 0.75
- For calculating power under the alternative hypothesis of non-inferiority, $A_2 = A_1$
- Correlation of A within reader, between reading conditions; r_1= 0.53
- Difference in correlation of A between reader, within reading condition versus between conditions; $r_2 - r_3$= 0.02
- Correlation between the set of A from each reading condition, r_b= 0.73
- Variance of A in population of readers, within reading condition, σ^2_b = 0.0016
- Variance of A when the same reader interprets the same case sample in the same reading condition, σ^2_w = 0.0001
- Variance of A because cases are a sample, σ^2_c, calculated using a binormal approximation (Obuchowski, 1994)[15]. This variance depends on the value of A and on the numbers of cancer and non-cancer cases in the sample.

Estimates of power for MRMC studies are highly dependent on the assumptions above. We therefore also obtained the number of readers required to provide 80% protected power allowing for possible attrition or parameter misspecification:

- 15% attrition rate for either cases or readers.
- Decrease in the comparator metric, A_1, to 90% of its assumed value.
- If power decreased when a parameter value increased, we increased said parameter by 50% on the measurement scale or by 125% on the variance scale.
- If power decreased when a parameter value decreased, we decreased said parameter by 33% on the measurement scale.

Table 8 shows that a study with **18** readers, **60** cancers, and **240** non-cancers, provides at least 80% protected power.

Table 9: Estimated Power for Primary Endpoint: Non-inferior AUC

	Power (%)*
Initial values**: A_1 = 0.80, δ= 0.05, *n* = 60 cancers + 240 non-cancers, σ^2_{ab} =0.0004 from σ^2_b = 0.0016 and r_b= 0.73, r_1 = 0.53, $(r_2 - r_3)$ = 0.02 and σ^2_w = 0.0001	92
$0.90A_1$	91
0.85n	89
0.85R	88
$2.25\sigma^{2ab}$	85
$0.67r_1$	88
$2.25\sigma^2_w$	91
$1.5(r_2 - r_3)$	89
*Power calculated using Student's *t* distribution with R – 1 degrees of freedom and rounded down to nearest whole percent.	
** A_1 = Area under the ROC curve with FFDM. δ = Non-inferiority margin. σ^2_{ab} = variance for interaction between reader and reading condition, obtained as the product of between-reader variance σ^2_b and 1 minus the correlation between the set of AUCs in the two reading conditions, r_b. r_1 = correlation of AUCs within reader, between reading conditions. $r_2 - r_3$ = difference in correlation of AUCs between reader, within reading condition versus between conditions. σ^2_w = within-reader variance. Variance because cases are a sample, σ^2_c, calculated using a binormal approximation (Obuchowski, 1994)[15]	

13 Sponsor and Investigational Site Requirements (Study Administration)

13.1 Ethics

13.1.1 Ethical Conduct of the Study

This study will be conducted in accordance with the ethical principles that have their origins in the Declaration of Helsinki.

A description of this clinical trial will be available on ***http://www.ClinicalTrials.gov***, as required by U.S. Law.

13.1.2 Institutional Review Board

The sponsor and investigator(s) shall assure that an IRB, constituted in accordance with 21CFR812 Subpart D and ICH guidelines for GCP (ICH E6 GCP) will provide initial and continuing review of the study.

Prior to enrollment of study subjects in the image acquisition portion of the study, documented IRB approval of the protocol, the informed consent form, and any other written materials supplied to the subjects, including any advertisements for subject recruitment, was obtained and provided to the sponsor or its designated representative(s).

Prior to commencement of the reader study portion of the study, documented IRB approval of the protocol, the reader informed consent form, and any other written materials supplied to the readers must be obtained by the sponsor or its designated representative(s).

The IRB must also be informed of any protocol amendments prior to implementation. The sponsor must provide reports of any change in research activity (e.g., the completion, termination, or discontinuation of the study) to the IRB. Annual review documentation will be submitted to the IRB 90 days prior to the IRB approval expiration.

13.2 Reader Informed Consent

Each reader must provide written informed consent before participating in the reader study of this protocol.

13.3 Protocol Compliance

Except for a change that is intended to eliminate an apparent immediate hazard to a study reader, the protocol shall be conducted as described. Any such change must be reported immediately to Fujifilm or its representative(s) and the governing IRB.

13.3.1 Protocol Amendments

Protocol amendments will be prepared and approved by Fujifilm or its authorized designee. All protocol amendments will be signed by the investigator and submitted to the IRB for review prior to implementation. Documentation of IRB approval must be provided to Fujifilm or its representative. If an amendment significantly alters the study design, increases potential risk to the study reader, or otherwise affects statements in the informed consent, the informed consent must be revised and submitted to the IRB for review and approval. The approved informed consent form must be used to obtain consent from new readers enrolled in the study and must also be used to re-consent readers already enrolled in the study, if they are affected by the amendment.

13.4 Retention of Study Records

All documents pertaining to the conduct of this clinical study – including eCRFs, informed consent forms, source documents, and other records must be retained during the investigation, and unless otherwise requested by the sponsor, for a period of 2 years after the later of the following two dates as per 21CFR812.140 (d):

- Records are no longer required for the purposes of supporting a regulatory submission to the FDA.

OR

- Following the termination or withdrawal of the regulatory submission.

The principal investigator must contact Fujifilm in writing prior to the destruction of any study records or in the event of accidental destruction or loss of the documents. If the principal investigator leaves the institution where the study is being conducted, the principal

investigator must contact the sponsor to arrange for transfer of responsibilities and alternative record storage options, if applicable.

Study documentation includes all eCRFs, data correction forms, source documents, monitoring logs, sponsor-investigator correspondence, protocols and amendments, ethics committee correspondence and approvals, signed consent forms, etc. Source documents include all original records or observations, results and activities necessary to reconstruct and evaluate the study. Source documents include but are not limited to tissue diagnosis reports, mammography images, subject progress notes, hospital charts, appointment books, radiology reports, and any other records or reports of procedures performed during the study.

13.5 Data Collection and Management

13.5.1 General

All data entry will be performed using an eDC system following the guidelines established in a Data Entry Guidelines document. An eDC Training & Registration form will be kept for all study site personnel which will include date of training, participant contact information, and eDC user privilege level. Training will include review of all study specific guidelines and data entry requirements.

The privacy of participating subjects must be maintained in accordance with the HIPAA privacy ruling and disclosed per 24CFR50.

The eCRFs will be data entered, and then reviewed for accuracy and completeness by a study monitor representing Fujifilm, as well as reviewed by DMC staff.

13.5.2 Data Management Center

The responsibilities of the DMC will include monitoring and oversight of the collection efforts, performance of the data audits, facilitating data entry into a computerized database, and providing the data to the statistician.

The DMC responsibilities will include oversight of the data collection efforts (excluding image data, which will be collected by Fujifilm or its designees), facilitation of data entry into the computerized database, all final data cleaning, and performance of a final data quality-control audit prior to database lock.

Electronic CRFs will be logged and tracked when received by the DMC. eCRFs will then be entered and stored in a 21 CFR Part 11 compliant clinical database. Upon completion of the study and resolution of all outstanding DCFs, the database will be locked and datasets necessary for analysis and reporting will be generated and provided to the statistician for final analysis.

13.5.3 New Suspicious Findings

If half or more of the study readers note a suspicious finding (BI-RADS 0, 3, 4 or 5) not documented by the image acquisition facility, the DMC will write a letter to the principal investigator at the image acquisition site noting the finding.

13.6 Confidentiality

13.6.1 Subject Confidentiality

To maintain subject confidentiality, no identifying patient data will be displayed. Unique image identifiers associated with the case ID number will be visible on each image for tracking purposes. A unique ID, one per case per modality (that is, a unique number for each DBT plus S-View case and a different unique number for each FFDM case) will be assigned to each subject whose images are selected for the reader study and will be the only subject identifier on all study-related documentation.

Prior to the initiation of the study, Fujifilm or its authorized representative will provide the readers with adequate training related to the identification. The subject's unique identifier may be displayed on the softcopy image as an overlay. The overlay may be turned on and off as necessary during the reader studies for reference.

13.6.2 Reader Confidentiality

Information collected during the pivotal reader study will be reported in such a way as to preclude identification of any individual radiologist's performance. A unique blinded ID number will be assigned to each radiologist participating in the pivotal reader study for entry on the eCRFs. A master list of these ID numbers for the radiologists participating in the pivotal reader study will be securely maintained by the Fujifilm project manager.

13.6.3 Transfer of Confidential Information to the Sponsor or the DMC

This reader study utilizes data from the Fujifilm library that was obtained in Fujifilm protocol FMSU2013-004A. All data utilized for this study (photocopies of radiology, pathology and clinical reports as well as images) was de-identified prior to removal from the acquisition center(s).

13.7 Study Monitoring

Fujifilm and/or its designees will monitor the study, verify the collection of data, and confirm that the study is being conducted according to the protocol. The eCRF data will be recorded.

The study may be audited by representatives from the FDA or other regulatory agencies, who also shall be allowed access to study documents.

If Fujifilm, the principal investigator, the IRB, or a regulatory authority discover conditions arising during the study that indicate that the study should be halted or that the study center's participation in the study should be terminated, this action may be taken after appropriate consultation between Fujifilm, its designee, and the investigator(s).

13.8 Sponsor Audits

Individuals appointed by the sponsor may visit the CRO to conduct an audit of the study GCPs. The purpose of the visit will be to determine adherence to the protocol, applicable regulations and the sponsor's procedures in addition to the assessing the accuracy of the study data (Inspection Readiness). Prior to initiating this audit, the CRO will be contacted by the sponsor to arrange a convenient time for this visit.

13.9 Inspection by a Regulatory Agency

At some point during the study, the investigator may be visited by a regulatory agency (e.g., FDA) to conduct an inspection of the study. The purpose of this visit will be to determine adherence to the protocol and regulatory requirements for conducting clinical studies. The investigator must immediately notify Fujifilm when contacted by a regulatory agency for the purposes of conducting an inspection of the Fujifilm's study.

13.10 Financial Disclosure

In accordance with 21CFR54, the principal investigator and the study readers will provide Fujifilm or its designee sufficient and accurate information on financial interests (proprietary or equity interests and payments exclusive of clinical study costs) to allow complete disclosure documenting lack of conflict of interest. The principal investigator, the pivotal study readers shall promptly update this information with any relevant changes that occur during the course of the study, at the completion of the study, and for a period of 1 year following the completion of the study.

13.11 Study or Study-site Termination

If Fujifilm, the principal investigator, the IRB, or a regulatory authority discover conditions arising during the conduct of the study that indicate that the study should be halted or that the study center should be closed, this action may be taken after appropriate consultation between Fujifilm and the principal investigator. Conditions that may warrant termination of the study include, but are not limited to, the following:

- A decision on the part of Fujifilm to suspend or discontinue testing, evaluation, or development of a product.

The study center may warrant closure/termination for the following reasons:

- Failure of the investigator to comply with pertinent regulatory authority regulations.
- Knowing submission of false information from the research center to Fuji, study monitor, or a regulatory authority.
- Insufficient adherence to the protocol requirements.

Study termination and follow-up will be performed in compliance with the conditions set forth in 21CFR812.150 (a) (2).

14 References

1. American Cancer Society. Cancer Facts & Figures. Available at: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2018.html. Accessed February 28, 2018.
2. American Cancer Society. Breast Cancer Risk and Prevention. Available at: https://www.cancer.org/cancer/breast-cancer/risk-and-prevention.html Accessed February 28, 2018.
3. Nelson HD, Tyne K, Naik A, et al. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Ann Intern Med.* 2009;151:727-737.
4. U.S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med.* 2009;151:716-726. Available at: http://www.ahrq.gov/clinic/uspstf09/breastcancer/brcanrs.pdf. Accessed February 3, 2009.
5. Hellquist BN, Duffy SW, Abdsaleh S, et al. Effectiveness of population-based service screening with mammography for women ages 40 to 49 years: evaluation of the Swedish Mammography Screening in Young Women (SCRY) cohort. *Cancer.* 2011 Feb 115; 117(4); 714-22. doi: 10.1002/cncr.25650
6. Sickles EA, Wolverton DE, Dee KE. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology.* 2002; 224:861-869.
7. Sickles EA, Miglioretti DL, Ballard-Barbash R, et al. Performance benchmarks for diagnostic mammography. *Radiology.* 2005; 235:775-790.
8. National Cancer Institute. Breast Cancer Surveillance Consortium data. Available at: http://breastscreening.cancer.gov/data/variables/. Accessed July 14, 2017.
9. Warner, E. Breast-Cancer Screening. *The New England Journal of Medicine.* 2011,365:1025-1032.
10. Conant E, Wan F, Thomas M, et al. Implementing Digital Breast Tomosynthesis (DBT) in a Screening Population: PPV1 as a Measure of Outcome. Presented at RSNA 2013, SSK01-02 Breast Imaging (Digital Breast Tomosynthesis Screening Outcomes)
11. Obuchowski NA, Rockette HE. Hypothesis testing of diagnostic accuracy for multiple readers and multiple tests: an ANOVA approach with dependent observations. Communications in Statistics - Simulations. 1995; 24(2): 285-308.
12. Obuchowski NA, on the comparison of correlated proportions for clustered data.Statistics in Medicine.1998 17(13): 1495–1507.
13. Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009 Feb 15;28(4):586-604.
14. Rao JNK, Scott AJ. A simple method for the analysis of clustered binary data, Biometrics 1992; 48:577–585.
15. Obuchowski NA. Computing sample size for receiver operating characteristic studies. Invest Radiol 1994; 29:238–243.
16. SenoClaire DBT Option to Senographe Essential FFDM System – Summary of Safety

and Effectiveness Data (SSED) for P130020 (August 26, 2014).

17. U.S. Department of Health and Human Services Food and Drug Administration CDER/CBER. Multiple Endpoints in Clinical Trials – Guidance for Industry. January 2017.
18. Selenia Dimensions 3D System – Summary of Safety and Effectiveness Data (SSED) for P080003 (September 24, 2010) and P080003/S001 (October 24, 2012).
19. Obuchowski RA, Lieber, ML, Powell, KA. Data analysis for detection and localization of multiple abnormalities with application to mammography. Acad. Radiol. 2000; 7: 516-525.

15 Appendices

Appendix A: Fujifilm FFDM and DBT Image Acquisition, Display, and Interpretation

Note: The proposed brand and/or configuration of the ASPIRE products are subject to change.

Overview of ASPIRE Cristalle and ASPIRE Bellus II

Fujifilm's **ASPIRE Cristalle** DBT Option with S-View (**Figure B**) mammography acquisition system generates both FFDM and DBT images. It is comprised of the X-ray exposure unit and acquisition workstation, which is FDA PMA approved (P160031) for mammography and digital breast tomosynthesis image acquisition. S-View images are created from the reconstructed DBT images. Without the need of additional radiation exposure to the patient, the S-View image is generated by "combining" the DBT images of that view into a single 2D synthetic mammography image. The S-View image, which may look similar to a FFDM image, is automatically labeled as "S-View" to avoid confusion with the FFDM image.

Fujifilm's **ASPIRE Bellus II** (**Figure C**) mammography review workstation displays and supports the physician's review of 2D mammography, DBT and S-View images. It is comprised of the display accessories of the ASPIRE Cristalle that include the major components of a PC (with image review and manipulation software installed), a console monitor, and two high-resolution grayscale diagnostic display monitors.

Figure B. ASPIRE Cristalle - Fujifilm's Mammography acquisition system

Figure C. ASPIRE Bellus II - Fujifilm's Mammography Viewer

Acquisition of FFDM and DBT plus S-View Images by ASPIRE Cristalle (Figure B)

All FFDM and DBT plus S-View images of the enrolled study subjects were acquired on the ASPIRE Cristalle system. For each study subject, the DBT and FFDM images were acquired under the same breast compression for each of the LCC, RCC, LMLO and RMLO views. The DBT images will be acquired first and then the x-ray tube will return to the normal in relation to the detector and the FFDM image will be acquired. The S-View images will be reconstructed by the acquisition workstation (AWS) of the ASPIRE Cristalle system.

Display of FFDM and DBT plus S-View Images on ASPIRE Bellus II (Figure C)

Mammography images are displayed on the ASPIRE Bellus II for the physician's review. The console allows the user to manage various aspects of the system configurations and setup, including the modality work list, local archive, communications, and the selection of images to be displayed on the two high-resolution grayscale diagnostic display monitors. Each CC and MLO view of the breast can be seen as a single FFDM image or a DBT plus S-View images. The viewer supports typical image-manipulation functions, such as windowing-and-leveling, zooming, etc.

Reading FFDM and DBT plus S-View Mammograms

Reading DBT plus S-View examination follows the traditional image-review process in FFDM examination except that the physician will read the DBT plus S-View images instead of a single FFDM image for each of the CC and MLO views of the breast. During the reading, the physician can control many aspects of the displayed images and hanging protocol(s).

Appendix B: Inclusion and Exclusion Criteria from FMSU2013-004A, Version 3.0

Study Population

All female subjects, of any ethnic or racial origin, who are appearing for a routine screening examination or have been referred for further diagnostic evaluation after a screening examination (within 60 days) or have a four-view mammogram (within 60 days) due to clinical concerns and receive a BI-RADS of 4 or 5 after diagnostic workup and are scheduled for biopsy will be eligible to participate in the study.

Subjects will receive a written explanation of the study and will be asked to provide written informed consent to participate. Once eligibility is confirmed and informed consent has been obtained, the subjects will be enrolled in the study.

Inclusion Criteria

Subjects enrolled must meet all the following inclusion criteria:

- Screening Subjects
  - Be at least 40 years of age, are
  - Asymptomatic,
  - Scheduled for a routine screening mammogram,
- Recall Subjects
  - Be at least 18 years of age
  - Received a BIRADS 0 within the last 60 days
  - Are recalled for additional imaging
- Diagnostic Subjects
  - Be at least 18 years of age,
  - Scheduled for a biopsy due to an assessment of BI-RADS® 4 or 5 after diagnostic work-up of a suspicious screening or clinical finding within the last 60 days.
- Have the ability to understand the requirements of the study, to provide written informed consent, and to comply with the study protocol, and
- Meet none of the exclusion criteria.

Exclusion Criteria

Subjects will be excluded from participating in the study if they meet any one or more of the following exclusion criteria:

- Presence of a breast implant.
- Women with only a single breast; for example, post mastectomy patients.
- Is pregnant or believes she may be pregnant.
- A woman who has delivered and who has expressed the intention to breast-feed or is currently breast-feeding.
- A woman who has significant existing breast trauma within the last one year.
- Has self-reported severe non-focal or bilateral breast pain affecting subject's ability to tolerate digital mammography and/or breast tomosynthesis examinations.
- A woman who has had a mammogram performed for the purpose of therapy portal planning within the last year.

- Cannot, for any known reason, undergo follow-up digital mammography and/or breast tomosynthesis examinations (where clinically indicated) at the participating institution.
- Is an inmate (see *US Code of Federal Regulations 45CFR46.306*)