White Paper

Superior Diagnostic Accuracy with Additional Digital Breast Tomosynthesis

Main findings of the clinical study for Premarket Approval of Mammomat Inspiration with Tomosynthesis Option

By Nancy Nalleweg, PhD, Thomas Mertelmeier, PhD,
Johannes Georg Korporaal, PhD, and Axel Hebecker, PhD

siemens.com/healthineers

SIEMENS Healthineers

Objective and setup

1 Objective and setup of the Premarket Approval study

The aim of this white paper is to summarize the clinical study that supported the Premarket Approval (PMA) of the "Mammomat Inspiration with Tomosynthesis Option" by the Food and Drug Administration (FDA) [1]. This clinical study contributed to the assessment of the safety and effectiveness of the device for breast cancer screening and diagnosis in the US. Further details of the study can be found in the Summary of Safety and Effectiveness Data released by the FDA [1]. This white paper also contains results from the clinical study beyond those published in the FDA document.

The objective of the clinical study was to demonstrate the superiority of 2-view Digital Breast Tomosynthesis (DBT) plus 2-view Full Field Digital Mammography (FFDM) to 2-view FFDM alone for the screening and diagnosis of breast cancer. The study, investigating the performance of 22 qualified radiologists (readers) in reviewing and scoring clinical images, consisted of a prospective case collection study to collect FFDM and DBT images and of a retrospective reader study. For the case collection study, a total of 800 subjects (corresponding to 800 cases) were enrolled from 7 United States clinical sites (Duke University, Durham, NC; SUNY, Stonybrook, NY; Brigham & Women's Hospital, Boston, MA; Cleveland Clinic, Cleveland, OH; NYU, New York, NY; St. Luke's Episcopal Hospital, Houston, TX; Miami Baptist Hospital, Miami, FL) between May 2011 and February 2014.

The reader study used a subset of 300 cases from this case collection study, meeting all criteria necessary for the study setup.

The cases were randomly selected from subgroups with specific mammographic characteristics: 165 negative cases, 85 biopsy-proven benign cases and 50 cancer cases (Figure 1). All enrolled subjects had 2-view FFDM and 2-view DBT images under mediolateral oblique (MLO) and craniocaudal (CC) positioning.

For the primary analysis[1] of the PMA study, the 300 subjects were reviewed on the breast level (490 breasts) and for the secondary analysis[2] on the subject level (300 subjects). 110 out of 600 breasts included in the reader study did not have 1-year follow-up information and were excluded from the primary analysis. Hence, 490 breasts corresponding to 347 normal breasts, 90 biopsy-proven benign breasts and 53 cancer breasts were investigated.

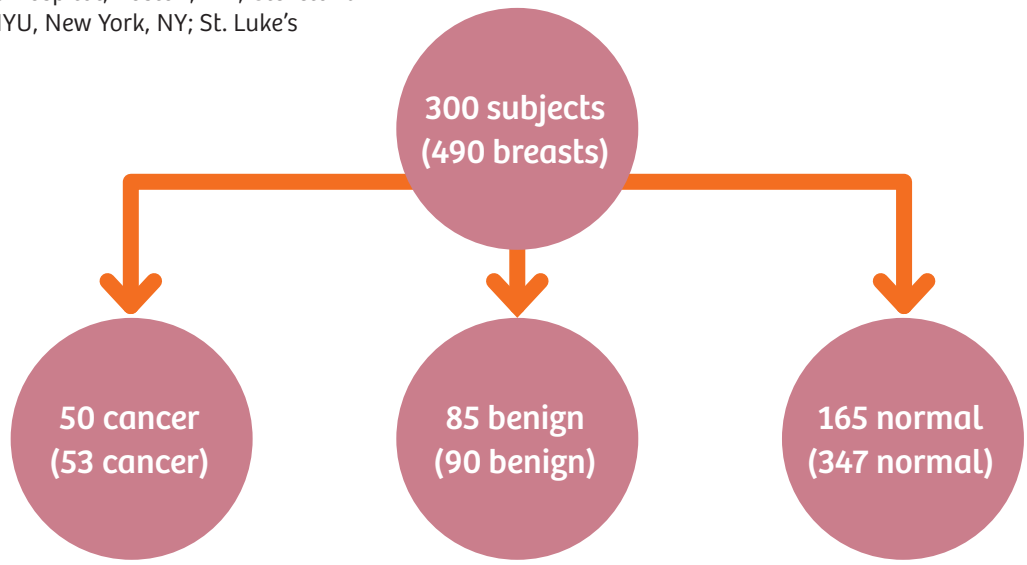


Figure 1: Study composition on the subject and breast level.

[1] During the primary analysis of a clinical study, the primary or most important question asked by the trial is investigated. The sample size of a clinical trial is powered for the primary analysis.
[2] During the secondary analysis, other relevant questions of the study are addressed and an additional post-hoc analysis is performed.

Clinical study results

2 Clinical study results

2.1 Improved accuracy through DBT for all readers and less interobserver variability

The accuracy of 22 radiologists in detecting and diagnosing malignant lesions was investigated for the "Mammomat Inspiration with Tomosynthesis Option". For this study, readers analyzed in one arm FFDM images alone and in the other arm FFDM images together with DBT scans. Figure 2 visualizes the improvement in readers' performance with the addition of DBT. Readers were sorted according to their area under the Receiver Operating Characteristic curve (AUC ROC) values in detecting and diagnosing malignant lesions with FFDM images alone (blue bars: from left (weaker performance) to right (stronger performance)) [2].

All 22 readers showed improvement in their diagnostic accuracy with the addition of 2-view DBT (blue plus red bars) (Figure 2). Secondly, weaker performing readers on the left side showed more improvement in diagnostic accuracy compared to readers on the right side who were already stronger with 2-view FFDM images alone. Furthermore, with the addition of 2-view DBT the differences in AUC ROC values between readers were less pronounced[3], showing a reduction in interobserver variability[4].

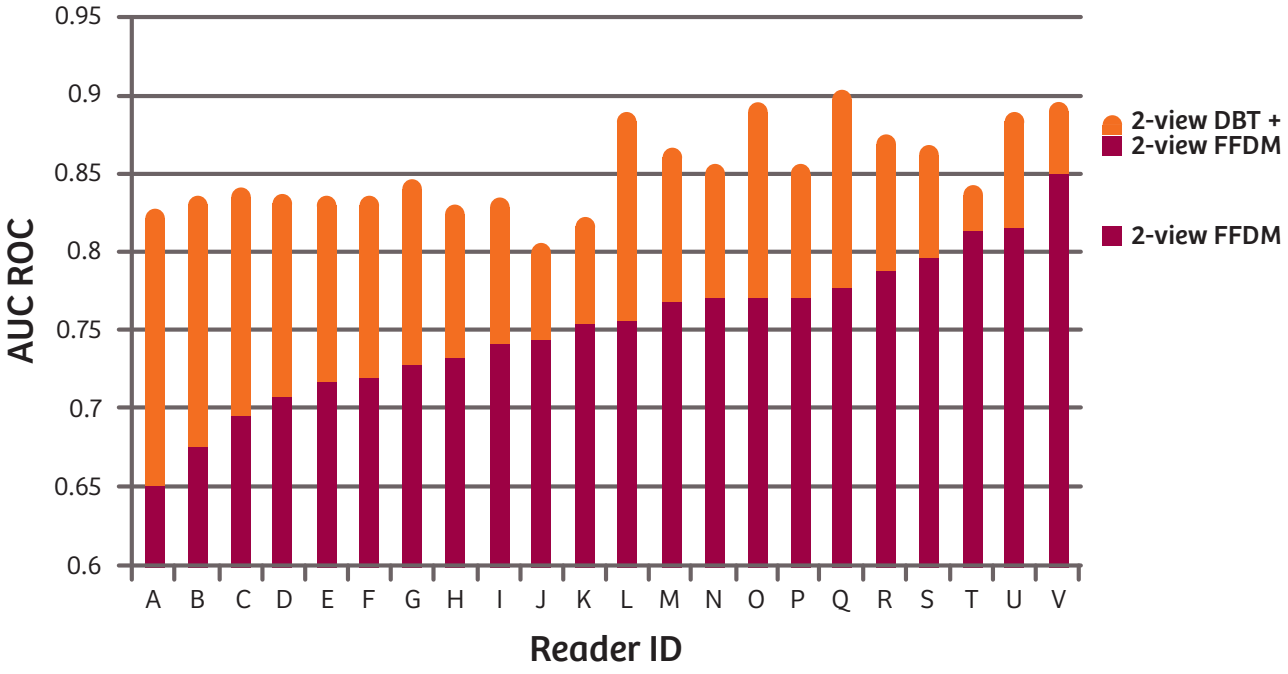


Figure 2: Readers' breast level AUC ROC with 2-view FFDM alone (blue) and 2-view FFDM plus 2-view DBT (blue plus red). Readers were sorted according to their accuracy in detecting and diagnosing malignant lesions with 2-view FFDM alone from left (lowest AUC ROC) to right (highest AUC ROC) [2].

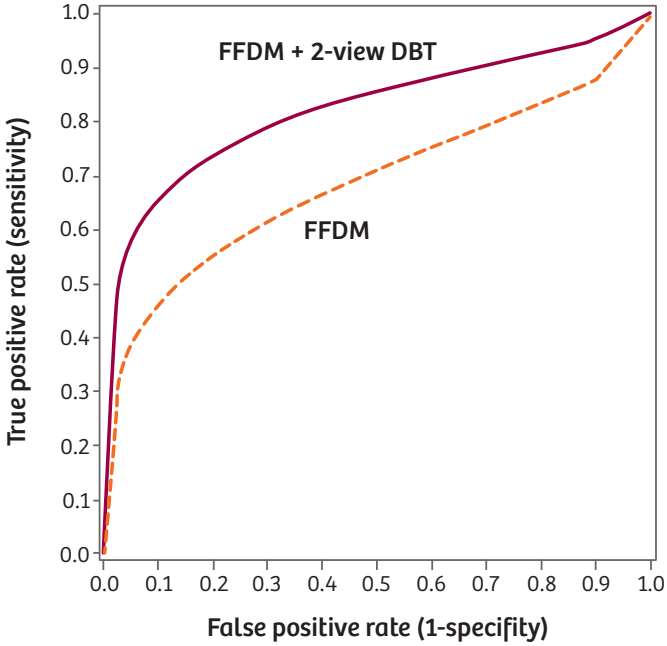
[3] 0.20 [2-view FFDM] vs. 0.10 [2-view FFDM + 2-view DBT] (calculated post-hoc)
[4] Indicated by the coefficient of variation: 6.4% [2-view FFDM] vs. 3.4% [2-view FFDM + 2-view DBT] (calculated post-hoc)

Clinical study results

2.2 Significant improvement in diagnostic accuracy with DBT

To assess the diagnostic accuracy of all readers, the average ROC curve of the 22 radiologists for 2-view FFDM plus 2-view DBT as well as for 2-view FFDM alone was plotted (Figure 3) [2]. The diagnostic accuracy in detecting and characterizing breast lesions significantly improved for all readers analyzing DBT images as an adjunct to FFDM images compared to analyzing FFDM images alone (Figure 3).

The average subject level AUC ROC was 0.709 (breast level: 0.752) for 2-view FFDM alone and 0.831 (breast level: 0.853) for 2-view FFDM plus 2-view DBT ($p < 0.0001$, subject level and breast level) [1] [2]. The readers demonstrated a 0.122 (breast level: 0.101) improvement in AUC ROC corresponding to an improvement of 12 (breast level: 10) percentage points and a relative improvement in diagnostic accuracy of 17% (breast level: 13%).



	Subject level	Breast level
2-view FFDM	0.709	0.752
2-view FFDM + 2-view DBT	0.831	0.853
Δ	0.122	0.101
Percentage points increase [%]	12	10
Relative increase [%]	17	13

Figure 3: Summary subject level ROC curve and subject/breast level AUC ROC values for 2-view FFDM alone and 2-view FFDM plus 2-view DBT for the 22 readers [1] [2].

Diagnostic accuracy improved by 17%

Clinical study results

2.3 Significant reduction in the non-cancer recall rate with DBT

The non-cancer recall rate was significantly reduced from 0.438 for 2-view FFDM alone to 0.355 for 2-view FFDM plus 2-view DBT (p = 0.0009) [1] [2]. This corresponds to a reduction of 19%. Thus, these subjects would not have to come back for a second appointment to receive unnecessary diagnostic follow-up procedures.

Non-cancer recall rate reduced by 19%

2.4 Increased sensitivity and specificity with DBT

The BI-RADS classification was used to score findings positive or negative. Patients with BI-RADS 3 findings during diagnostic workup either have to come back for a biopsy or are sent for surveillance, while subjects classified as BIRADS 1 and 2 in either screening or workup will not be requested to come back for follow-up. The addition of 2-view DBT improved the sensitivity by 0.191 (relative increase of 32%) when BI-RADS 4, 5 were counted as a positive test and by 0.150 (relative increase of 22%) when BI-RADS 3, 4, 5 were counted as a positive test (Table 1) [1]. An increase in specificity could only be observed when BI-RADS 3, 4, 5 (0.031 improvement) were counted as a positive test (Table 1) [1].

	2-view FFDM	2-view FFDM + 2-view DBT	Δ	Percentage points increase [%]	Relative increase [%]
Sensitivity with BI-RADS 4, 5 as a positive test					
Subject level	0.595	0.786	0.191	19	32
Specificity with BI-RADS 4, 5 as a positive test					
Subject level	0.732	0.730	-0.002	0	0
Sensitivity with BI-RADS 3, 4, 5 as a positive test					
Subject level	0.672	0.822	0.150	15	22
Specificity with BI-RADS 3, 4, 5 as a positive test					
Subject level	0.562	0.593	0.031	3	6

Table 1: Sensitivity and specificity observed in the study [1].

Sensitivity increased by 32% (BI-RADS 4-5) and 22% (BI-RADS 3-5)

Clinical study results

2.5 Improved diagnostic accuracy for different breast and lesion types with DBT

The mean diagnostic accuracy for different breast compositions (dense and fatty breasts) as well as lesion types (microcalcifications and masses) did improve with 2-view FFDM plus 2-view DBT compared to 2-view FFDM alone (Table 2). For dense breasts, AUC ROC increased from 0.734 for 2-view FFDM alone to 0.851 for 2-view FFDM plus 2-view DBT [1]. Thus, the readers demonstrated a 0.117 improvement in diagnostic accuracy corresponding to an improvement of 12 percentage points and a relative improvement of 16%.

Overall, there was an improvement in the diagnostic accuracy for all investigated breast and lesion types with the addition of 2-view DBT.

	2-view FFDM	2-view FFDM + 2-view DBT	Δ	Percentage points increase [%]	Relative increase [%]
Dense breast	0.734	0.851	0.117	12	16
Fatty breast	0.787	0.870	0.083	8	11
Masses	0.782	0.891	0.109	11	14
Microcalcification	0.744	0.790	0.046	5	6

Table 2: Breast level AUC ROC stratified by breast density and lesion type [1].

Diagnostic accuracy improved for all investigated breast types (dense, fatty) and lesion types (masses, microcalcifications)

2.6 Improvement in diagnostic accuracy with 1-view DBT (MLO)

For the 22 readers, the mean diagnostic accuracy in detecting and characterizing breast lesions already improved with the addition of only 1-view DBT (MLO) as an adjunct to 2-view FFDM. In this case, the subject level AUC ROC increased from 0.709 for 2-view FFDM to 0.804 for 2-view FFDM plus 1-view DBT (MLO) [1]. This 0.095 increase in diagnostic accuracy is an improvement of 10 percentage points and a relative improvement of 13%, which is consistent with other studies where an increase in the detection rate for 1-view DBT (MLO) alone was observed compared to 2-view FFDM [3] indicating a potential for dose saving.

Diagnostic accuracy improved by 13% with 1-view DBT (MLO)

3 Summary

The clinical study described in this paper supported the PMA of the "Mammomat Inspiration with Tomosynthesis Option" by the FDA. It shows the superiority of 2-view DBT plus 2-view FFDM to 2-view FFDM alone for the screening and diagnosis of breast cancer.

All readers significantly improved their diagnostic accuracy when analyzing DBT images as an adjunct to FFDM images compared to analyzing FFDM images alone.

Weaker performing readers improved their diagnostic accuracy to a higher extent compared to readers who were already stronger with 2-view FFDM images alone. The differences in diagnostic accuracy between readers were less pronounced, showing a reduction in interobserver variability with the addition of 2-view DBT.

The non-cancer recall rate was significantly reduced when analyzing DBT images as an adjunct to FFDM images compared to analyzing FFDM images alone. Thus, with the addition of DBT, fewer women will receive unnecessary diagnostic follow-up procedures.

The sensitivity and specificity increased through the addition of 2-view DBT to 2-view FFDM.

The diagnostic accuracy for all investigated breast types (dense, fatty) and lesion types (masses, microcalcifications) improved with the addition of 2-view DBT.

The diagnostic accuracy even improved for all readers with the addition of 1-view DBT (MLO) to 2-view FFDM compared to 2-view FFDM alone.

Appendix

4 Appendix: Important terminology

2D FFDM and 3D DBT

For both screening and diagnostic mammography, **Full Field Digital Mammography (FFDM)** provides 2D images of the breast. These 2D images can be taken from different angles, with two views commonly used for each breast, being the mediolateral oblique (MLO) and craniocaudal (CC) view. This results in the term "2-view FFDM".

Digital Breast Tomosynthesis (DBT), a 3D imaging technique [4], uses a number of low-dose exposures while moving the X-ray source over an arc above the breast. The set of projections is then used to reconstruct a 3D data set of the breast. DBT images are typically taken from the MLO view or from both the MLO and the CC view, referred to as "1-view DBT (MLO)" and "2-view DBT".

Sensitivity and specificity

For any diagnostic test, the results can be divided into those who get a positive and those who get a negative test result (Figure 4).

Subjects who get a positive test result are either true positives (TP, ill people correctly identified as such) or false positives (FP, healthy people falsely identified as being ill). Subjects who receive a negative test result either belong to the true negatives (TN, healthy people correctly identified as such) or the false negatives (FN, ill people falsely identified as being healthy).

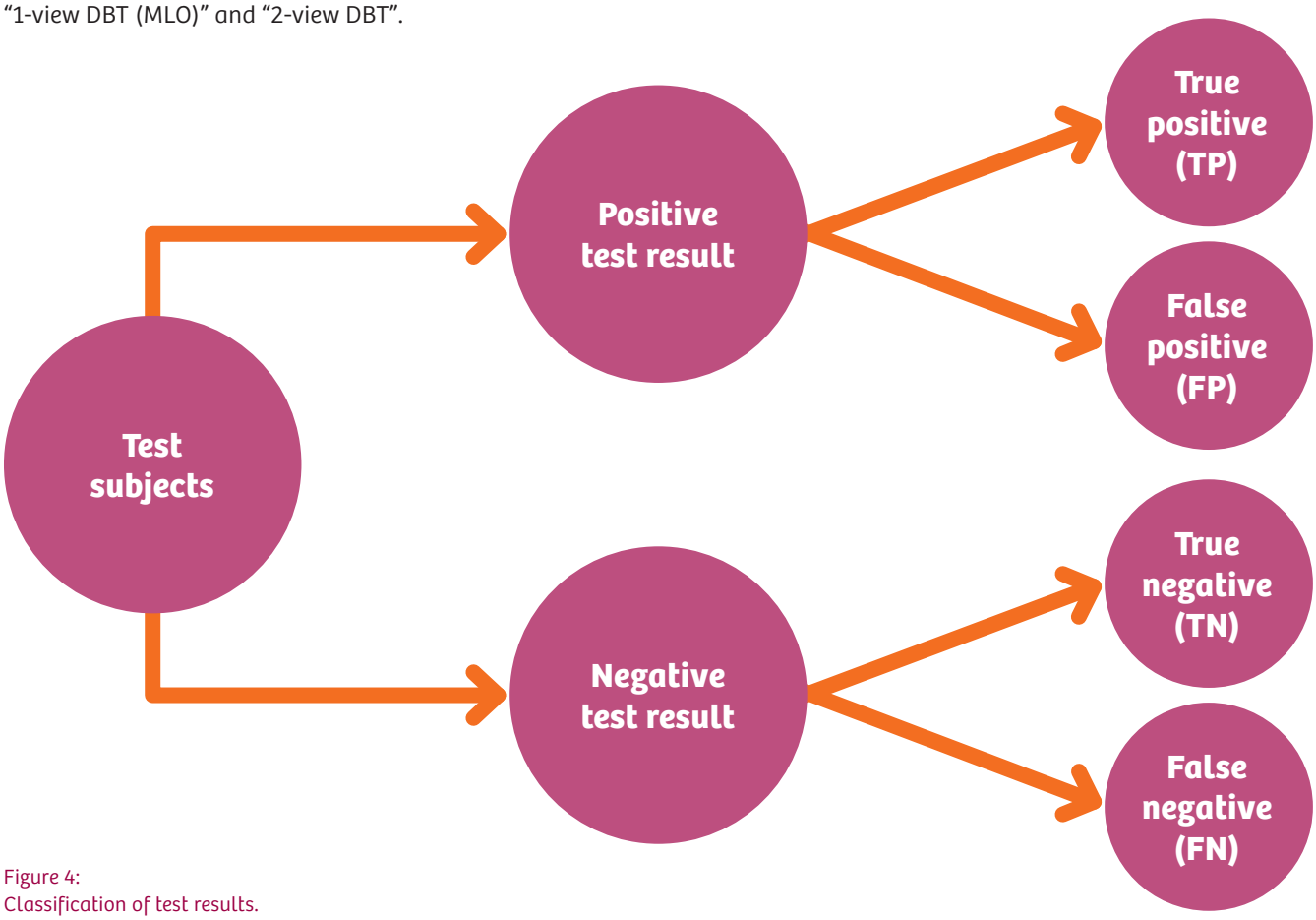


Figure 4:
Classification of test results.

Appendix

Sensitivity and specificity are statistical values often used in clinical studies.

Sensitivity is the proportion of truly diseased persons in the screened population who are identified as diseased by the screening test [5]. Sensitivity is also called the **true positive rate:**

$$\text{Sensitivity} = \text{True positive rate} = \frac{TP}{TP + FN}$$

Sensitivity:
Proportion of truly diseased persons in the screened population who are identified as diseased by the screening test

A high sensitivity corresponds to a high number of true positives and few false negatives which is highly desirable in screening and diagnostic mammography.

Specificity corresponds to the proportion of truly non-diseased persons in the screened population who are identified as non-diseased by the screening test [5].

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Specificity:
Proportion of truly non-diseased persons in the screened population who are identified as non-diseased by the screening test

The term "1-specificity" corresponds to the **false positive rate:**

$$\text{False positive rate} = \text{1-Specificity} = \frac{FP}{TN + FP}$$

A high specificity corresponds to few false positive results and is highly desirable because it helps to minimize the number of subjects receiving unnecessary diagnostic procedures. Both sensitivity and specificity (as well as true positive rate and false positive rate) can obtain numerical values between 0 and 1, corresponding to 0% and 100% respectively.

Sensitivity and specificity are more expressive than the **non-cancer recall rate** which corresponds to the number of subjects recalled for further assessment but finally diagnosed to have no cancer (FP) as a proportion of all subjects who had a screening examination.

$$\text{Non-cancer recall rate} = \frac{FP}{\text{Screening population}}$$

Thus, the non-cancer recall rate allows a statement to be made about the number of subjects recalled for unnecessary diagnostic follow-up procedures but it does not provide information about the proportion of subjects correctly identified by the test as having or not having the disease.

Appendix

AUC ROC as a measure of test accuracy

The Receiver Operating Characteristic (ROC) curve is a graphical representation of the relation between the true positive rate (sensitivity) and the false positive rate (1-specificity) (Figure 5). Each point on the ROC curve displays a true positive rate/false positive rate pair revealing the compromise between sensitivity and specificity, as an increase in sensitivity is characterized by a decrease in specificity.

The ROC curve for a perfect test matches the left-hand border and the top border (green line) whereas a test that matches the diagonal (purple line) is the same as random choice. The points on curve (1) and (2) in Figure 5 have the same sensitivity (true positive rate = 0.8), however, point (2) has a higher false positive rate than point (1) and therefore lower specificity than point (1). The accuracy of the test is measured by the area under the ROC curve (AUC ROC). Thus, for the two ROC curves in Figure 5, test (1) is more accurate than test (2). The value of AUC ROC will be between 1 for a perfect test and 0.5 being the equivalent of a random guess.

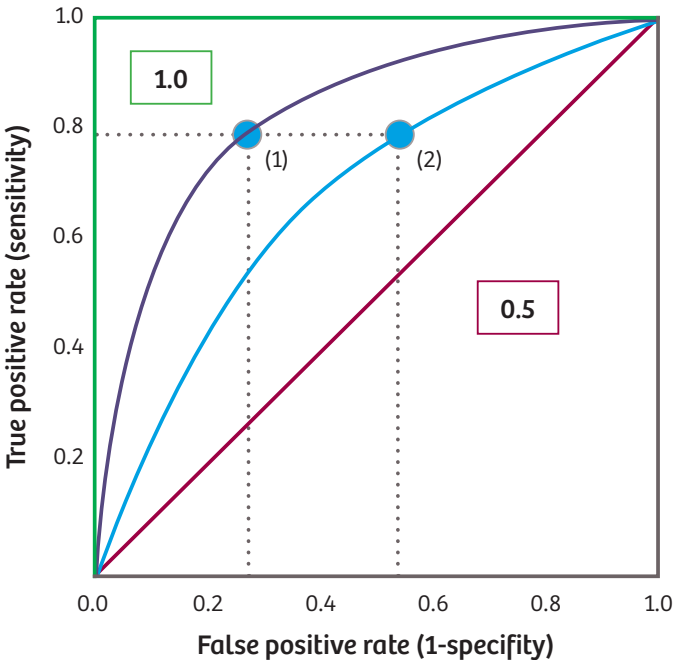


Figure 5: ROC curves with different values for the AUC ROC.

ROC (Receiver Operating Characteristic) – Relation between the true and false positive rate

AUC ROC (Area under the ROC curve) – Measure of the diagnostic accuracy of the test

- **AUC ROC = 1 → Perfect test**
- **AUC ROC = 0.5 → Random guess**

Glossary

Abbreviations

ACA	Area under the ROC curve
CC	Craniocaudal
DBT	Digital Breast Tomosynthesis
FDA	Food and Drug Administration
FFDM	Full Field Digital Mammography
FN	False negative
FP	False positive
MLO	Mediolateral oblique
PMA	Premarket Approval
ROC	Receiver Operating Characteristic
TN	True negative
TP	True positive

Glossary of terms

FFDM	2D images of the breast
DBT	3D volume created from a set of projections moving the X-ray source over an arc above the breast
2-view	MLO view plus CC view
1-view	MLO view or CC view

References

[1] Siemens Medical Solutions USA, Inc., "PMA P140011: FDA Summary of Safety and Effectiveness Data," FDA, 2015.

[2] Siemens Medical Solutions USA, Inc., "Clinical Study Report for Protocol SMS-SP09-01," Siemens, 2014.

[3] K. Lång, I. Andersson, A. Rosso, A. Tingberg, P. Timberg and S. Zackrisson, "Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmö Breast Tomosynthesis Screening Trial, a populationbased study," European Radiology, vol. 26, no. 1, pp. 184-90, 2016.

[4] J. Barkhausen, A. Rody and F. K. Schäfer, Digital Breast Tomosynthesis, Stuttgart/New York: Georg Thieme Verlag, 2016.

[5] N. Perry, M. Broeders, C. de Wolf, S. Törnberg, R. Holland, L. von Karsa and E. Puthaar, "European guidelines for quality assurance in breast cancer screening and diagnosis," European Commission, Luxembourg, 4th edition, 2006.

Siemens Healthineers Headquarters
Siemens Healthcare GmbH
Henkestr. 127
91052 Erlangen
Germany
Phone: +49 9131 84 0
siemens.com/healthineers