Clinical performance of digital breast tomosynthesis compared to digital mammography: blinded multi-reader study

G. Gennaro[(1)], A. Toledano[(2)], E. Baldan[(1)], E. Bezzon[(1)], C. di Maggio[(1)], M. La Grassa[(1)], L. Pescarini[(1)], I. Polico[(1)], A. Proietti[(1)], A. Toffoli[(1)]

[(1)] Oncological Institute of Veneto, I.R.C.C.S., Padova - Italy
[(2)] Biostatistics Consulting, LLC, Toronto - Canada

Mammography limitations

SCREENING MAMMOGRAPHY = BEST METHOD FOR EARLY DETECTION OF BREAST CANCER

FALSE NEGATIVE RATE = 10%-30%

Retrospective studies = 67% of cancers were visible on the prior mammograms

MISSED CANCERS → **DETECTION ERRORS**; **CHARACTERIZATION ERRORS**

Inter-observer variability

k statistics

0.00-0.20	SLIGHT
0.21-0.40	FAIR
0.41-0.60	MODERATE
0.61-0.80	SUBSTANTIAL
0.81-1.00	ALMOST PERFECT

Table 4

Interobserver Variability in Assigning Final BI-RADS Assessment Category

BI-RADS Category	κ Value
2	0.27
3	0.32
4a	0.14
4b	0.16
4c	0.26
5	0.56
Combined*	0.28

* BI-RADS categories 2–5.

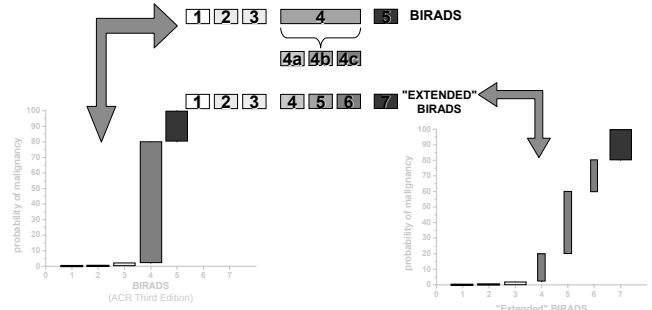
Lazarus et al
BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value
Radiology 2006, 239:385-391

5 radiologists

BI-RADS classification

BIRADS 3rd EDITION :

- **5 classes, increasing with probability of malignancy;**
- **BIRADS 4 (suspicious) includes most of probability distribution (5-80 %).**



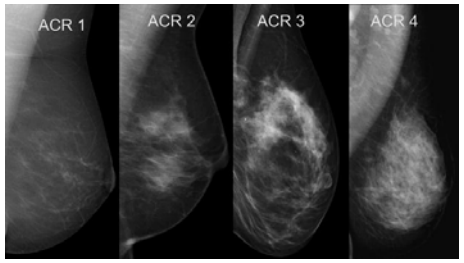
Classification accuracy can be improved by dividing BIRADS 4 category into 3 subclasses, corresponding to different risks of malignancy.

BIRADS 4th EDIITON:

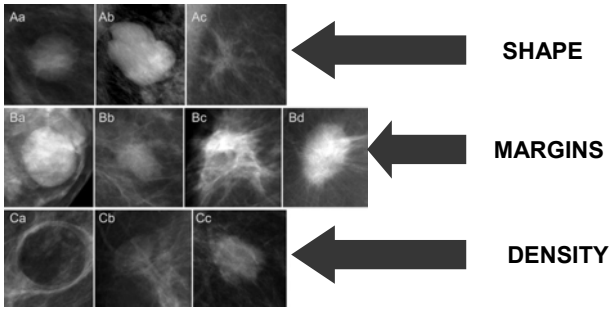
- **7 classes, increasing with probability of malignancy**

BI-RADS: breast density

ACR	Description	Diagnostic accuracy
1	Mostly fatty	Very high
2	Fibroglandular	High
3	Heterogeneously dense	Limited
4	Dense	Limited



BI-RADS: masses

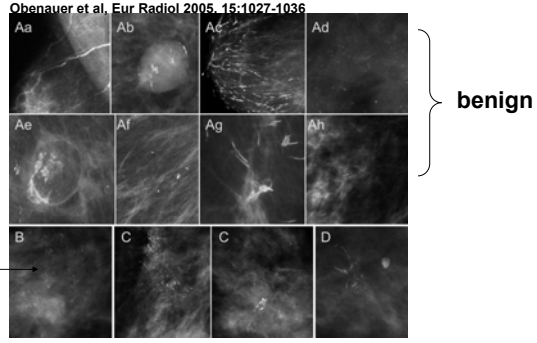


Obenauer et al, Eur Radiol 2005, 15:1027-1036

A LESION CAN BE CONSIDERED A MASS IF IT IS VISIBLE IN BOTH CC AND MLO VIEWS

BI-RADS: calcifications

Obenauer et al, Eur Radiol 2005, 15:1027-1036

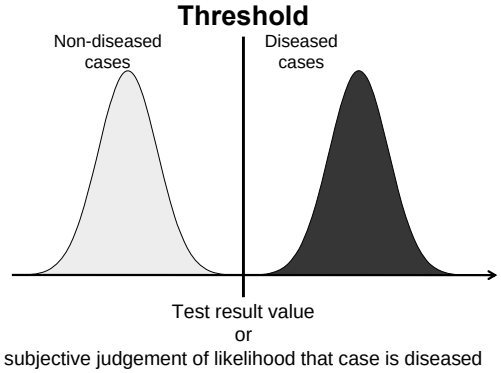


benign

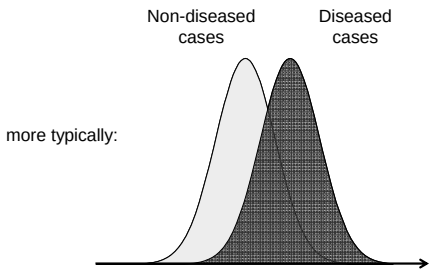
concern

high probability of malignancy

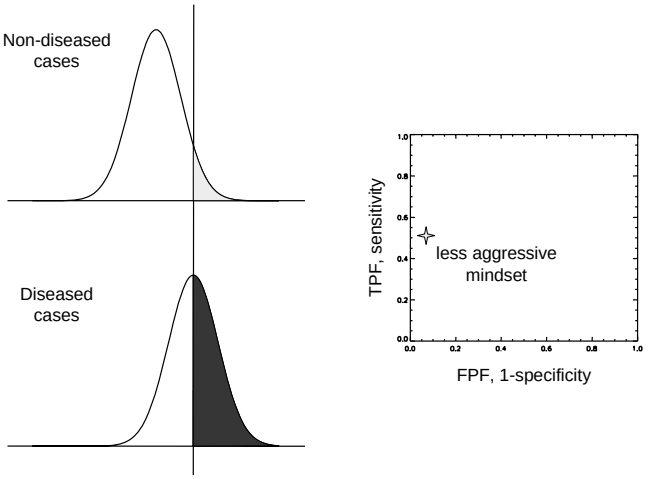
The ROC paradigm

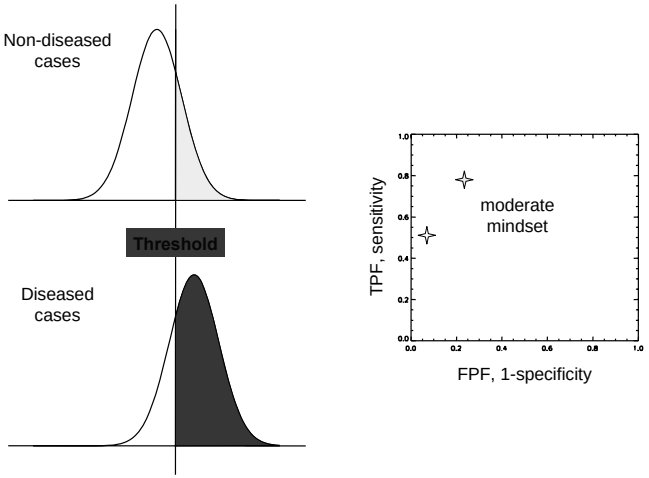


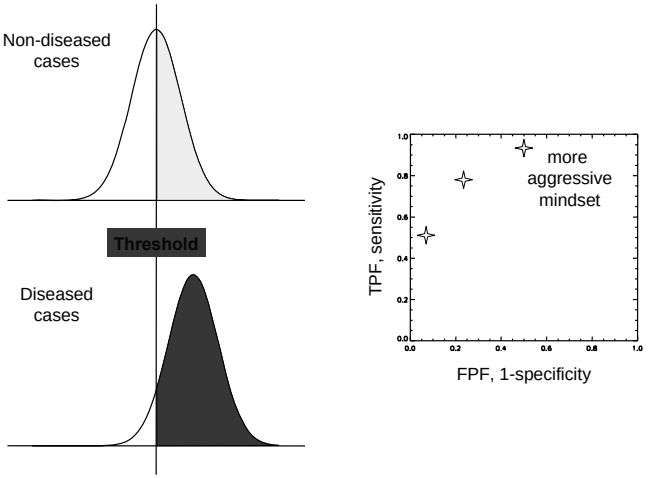
The ROC paradigm

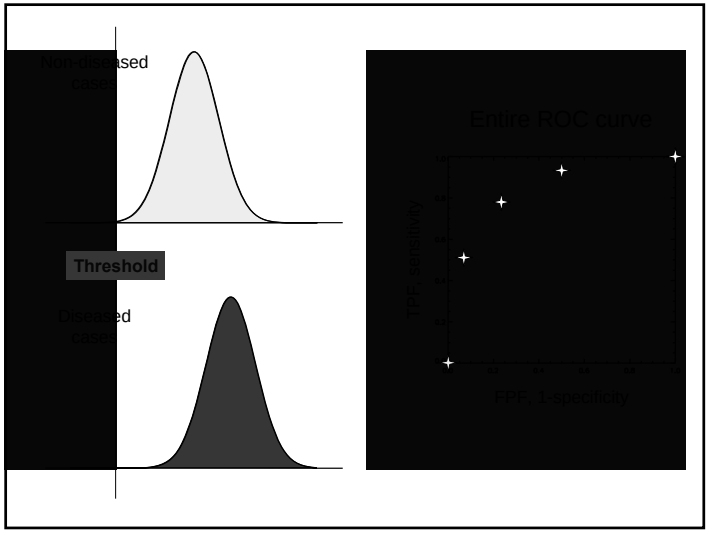


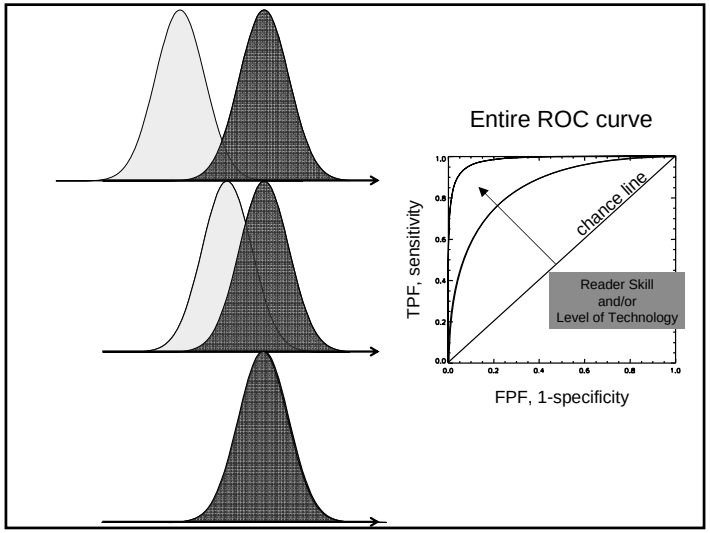
Test result value
or
subjective judgement of likelihood that case is diseased







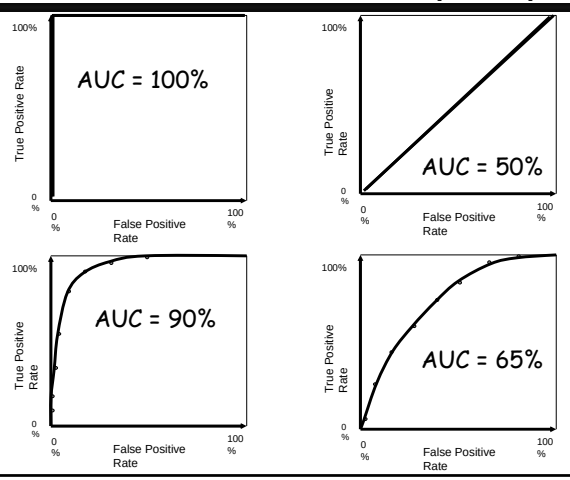




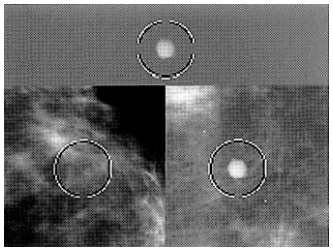
Area under ROC curve (AUC)

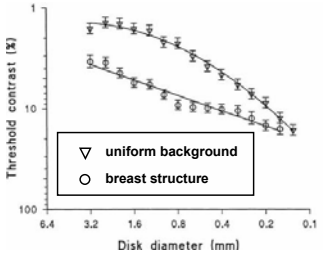
- *Overall measure* of test performance
- *Comparisons* between two tests based on differences between (estimated) AUC

Area under ROC curve (AUC)



Background Anatomic Noise





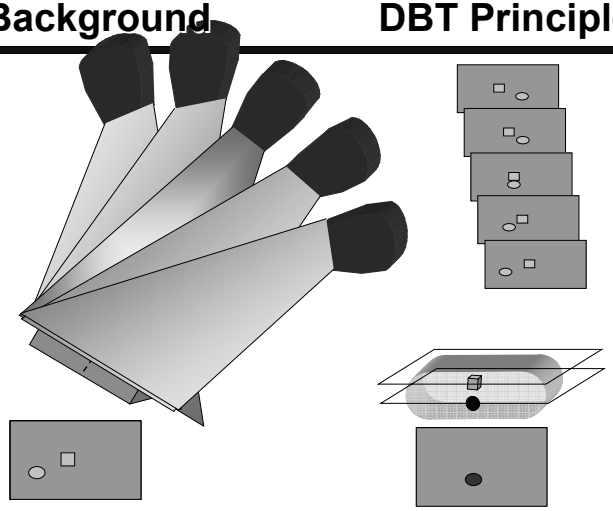
Burgess A E et al
Lesion detection in digital mammograms
Proc SPIEE 2001, 4321:555-560

Kotre C J
The effect of background structure on the detection of low contrast objects in mammography
BJR 1998, 71:1162-67

< LESION DETECTABILITY

< CLINICAL PERFORMANCE (sensitivity, specificity)

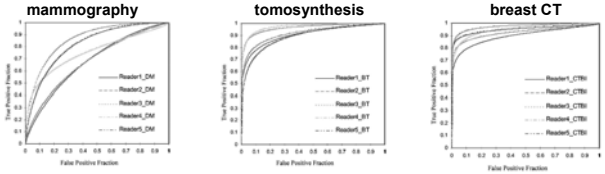
Background DBT Principles

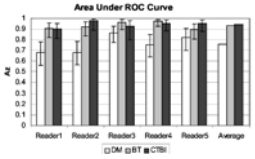


DBT & CTBI: accuracy

Gong X et al
A computer simulation study comparing lesion detection accuracy with digital mammography, breast tomosynthesis, anc cone-beam CT breast imaging
Med Phys 2006, 33:1041-1052

digital mammography | **tomosynthesis** | **breast CT**





DETECTION:SIGNIFICANT BENEFIT

Potential of DBT & CTBI

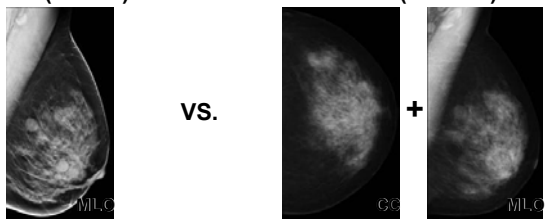
WILL DIGITAL BREAST TOMOSYNTHESIS [OR BREAST CT] REPLACE SCREENING MAMMOGRAPHY ?

[Dr. Dan Kopans, MGH]

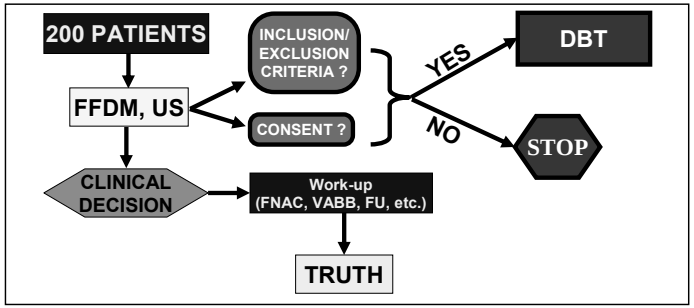
Purpose

COMPARE CLINICAL PERFORMANCE OF ONE-VIEW DIGITAL BREAST TOMOSYNTHESIS (DBT) VERSUS TWO-VIEWS FULL-FIELD DIGITAL MAMMOGRAPHY (FFDM)

DBT (1 view) **FFDM (2 views)**



Method: study population



INCLUSION CRITERIA

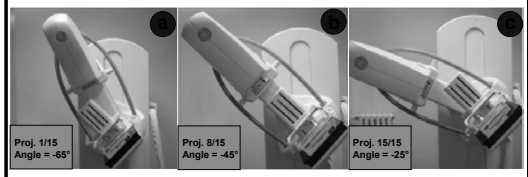
- >= 40 y
- lesion BIRADS >=3 (FFDM or US)
- breast size to fit detector FOV (19 x 23 cm2)

EXCLUSION CRITERIA

- previous breast mastectomy
- breast implant
- high genetic risk

Acquisition systems

DBT



**GE DBT investigational device*

CsI/a-Si flat panel;
- **19 x 23 cm²;**
- **100x100 µm² pixel size;**
- **Mo/Mo, Mo/Rh, Rh/Rh;**
- **Manual exposure mode;**
- **15 projections per breast;**
- **40° arc;**
- **MLO only.**

FFDM

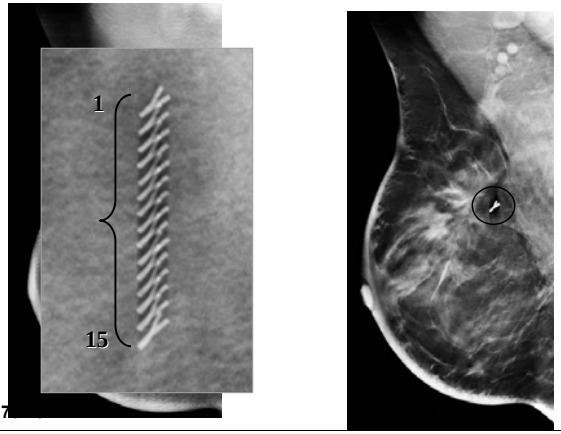
GE Senographe 2000D

CsI/a-Si flat panel;
- **19 x 23 cm²;**
- **100x100 µm² pixel size;**
- **Mo/Mo, Mo/Rh, Rh/Rh;**
- **AOP/STD**
- **CC + MLO**

Case # 69 - DBT projections

FFDM **DBT**

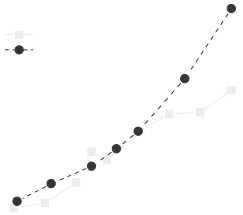
RMLO **RMLO**



slice 7

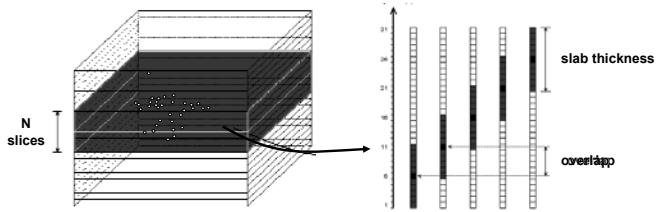
DBT dose

DBT Dose$_{MLO}$< = SFM Dose$_{CC+MLO}$

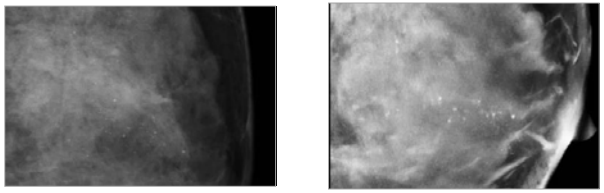


DBT dose (MLO view) per each breast thickness is compared with 2x Dose Acceptance Limits proposed by the "European Guidelines for quality assurance in breast cancer screening and diagnosis, 4th edition", derived from screen/film mammography

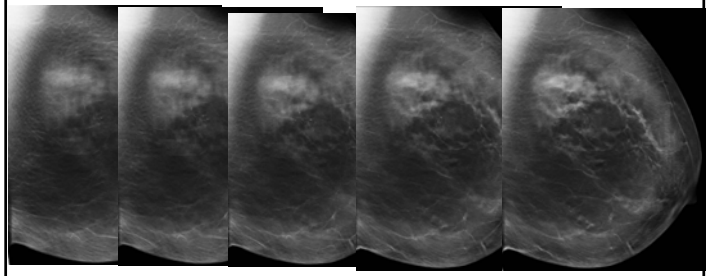
DBT reconstruction: slabs vs. slices



1 SLAB = 10 SLICES



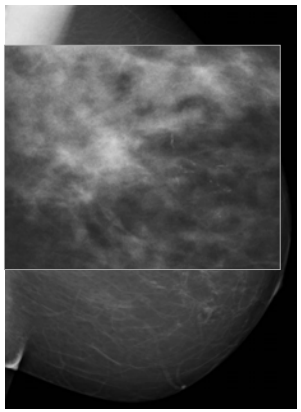
Case # 12 - DBT



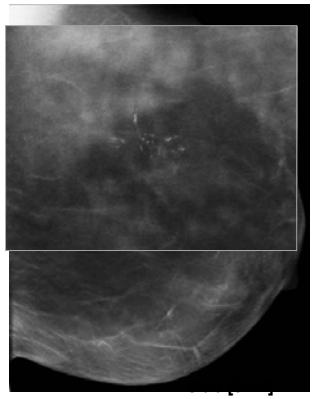
slice 8 **slice 9** **slice 10** **slice 11** **slice 12**

Case # 12 - FFDM vs DBT (slabs)

FFDM
LMLO

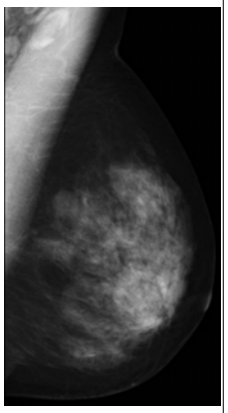


DBT
LMLO

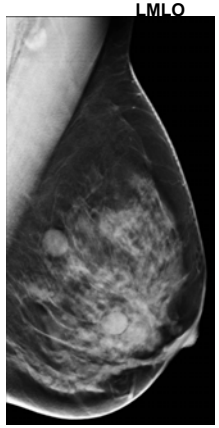


Case # 81 - FFDM vs DBT

FFDM
LMLO

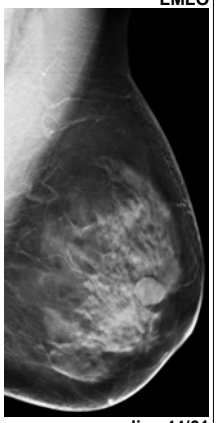


DBT
LMLO



slice 25/61

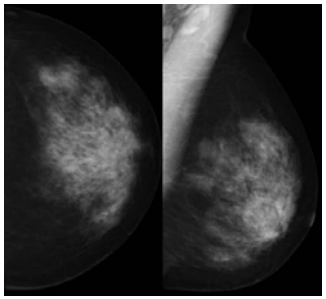
LMLO

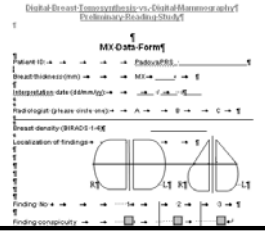


slice 44/61

Method: Reading Protocol

FFDM: CC+MLO



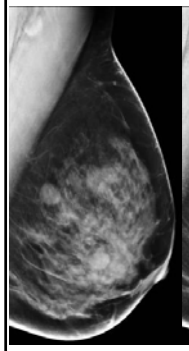


BIRADS (7 steps)

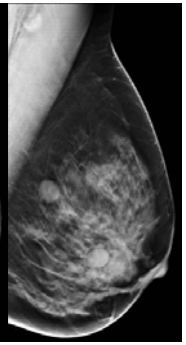
TRUTH

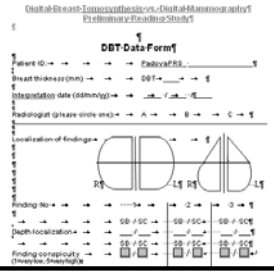
Method: Reading Protocol

DBT slabs

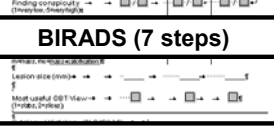


DBT slices





BIRADS (7 steps)



TRUTH

Method: Reading Protocol

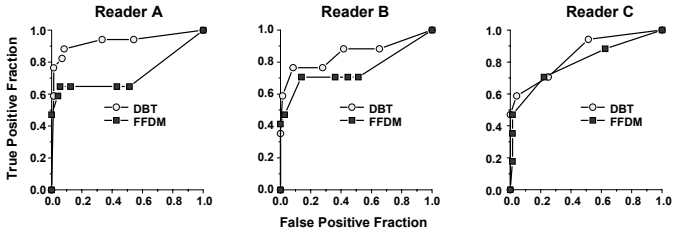
	FFDM		DBT slabs	DBT slices
Breast density (BIRADS 1-4)	X			
Localization of findings (max 3)	X		X	X
Finding Conspicuity (1-5)	X		X	X
BIRADS (7-steps)	X		X	X
Lesion type	X		X	X
Lesion size (mm)	X		X	X
Most useful view	CC	MLO	SLABS	SLICES

Methods Truth establishment

- ✓ Malignant lesions: histology (from surgical or core biopsy);
- ✓ Benign lesions: histology (in case of biopsy), FNAC and/or long/term follow-up (long-term >= 1y history).
- ✓ Negative cases (no lesion): information from the patient folder or consensus meeting (in case of disagreement)

First results: clinical performance

Readers 3
Patients 50
Effective cases 90



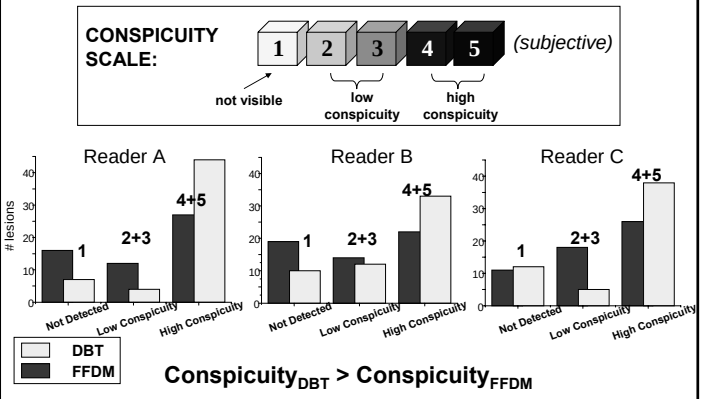
Reader A: AUC_{DBT} = 0.929, AUC_{FFDM} = 0.728, p = 0.003

Reader B: AUC_{DBT} = 0.852, AUC_{FFDM} = 0.757, p = 0.229

Reader C: AUC_{DBT} = 0.846, AUC_{FFDM} = 0.797, p = 0.522

Presented at ECR 2008

First results: image quality



$Conspicuity_{DBT}$ > $Conspicuity_{FFDM}$

Presented at IWDM 2008

First conclusions

1. $Conspicuity_{DBT}$ > $Conspicuity_{FFDM}$
2. AUC_{DBT} > AUC_{FFDM}
3. Significant difference for 1 of 3 readers
4. Sample size!!!

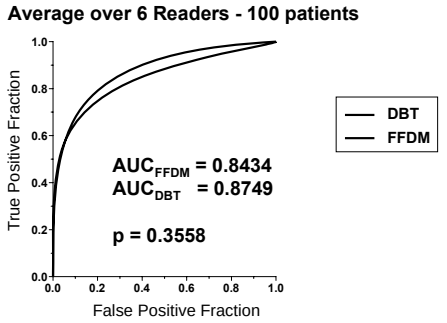
RESULTS ARE ENCOURAGING AND SUPPORT THE POTENTIAL BENEFIT OF TOMOSYNTHESIS OVER 2D-MAMMOGRAPHY

Second step

Readers	**3**	→	**6**
Patients	**50**	→	**100**
Effective cases	**90**	→	**197**

SAME READING PROTOCOL

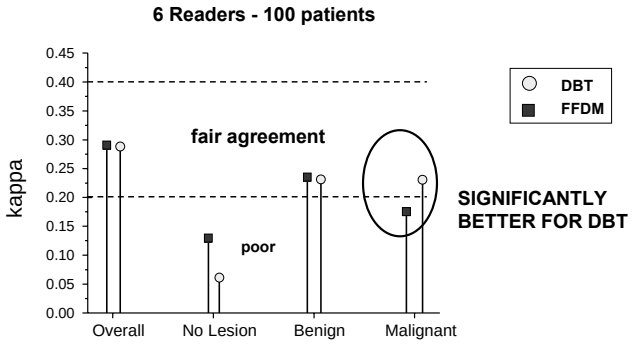
Clinical performance: MRMC ROC



NOT SIGNIFICANTLY DIFFERENT AT 95% CL

Presented at RSNA2008

Inter-reader variability



SIGNIFICANTLY BETTER FOR DBT

Presented at RSNA2008

Second conclusions...

1. Clinical performance of DBT (MLO) was slightly superior vs. FFDM (CC+MLO), even if not statistically significant;

2. Inter-reader variability was lower with DBT vs. FFDM for malignant lesions.

RESULTS SUPPORT THE OPPORTUNITY FOR TOMOSYNTHESIS TO REDUCE INTER-READER VARIABILITY IN AREAS UNDER ROC CURVES AND IN BIRADS SCORES FOR MALIGNANT LESIONS

Third step (final)

Readers	**6**	**6**
Patients	**100**	**200**
Effective cases	**197**	**371**

SAME READING PROTOCOL

Final results presented at

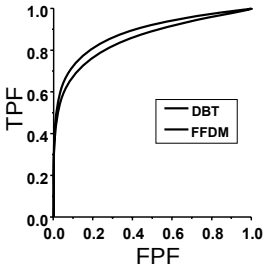
Method: Multi-Reader Multi-Case ROC

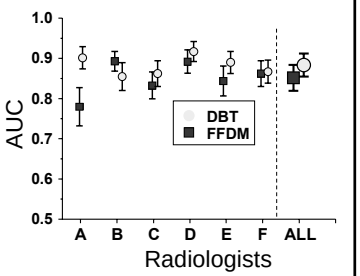
- Multiple Readers: 6 breast radiologists (5-30 y experience)
- Population: 200 patients
- Independent readings of left & right breasts = 371 effective cases
- Multiple Reading Sessions: including 50% DBT & 50%FFDM images
- Bias Control: NO DBT&FFDM images of the same breast in the same session

Results: MRMC ROC analysis

Malignant lesions vs. all other breasts

Average over 6 Readers





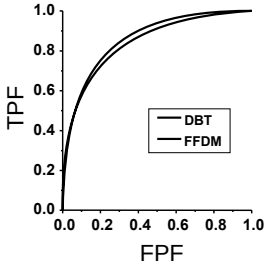
NOT SIGNIFICANTLY DIFFERENT AT 95% CL

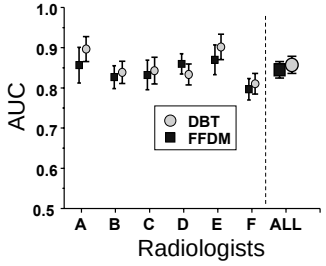
VARIABILITY IN AUC's ACROSS READERS SLIGHTLY LOWER FOR DBT (SD=0.0247) VS. FOR FFDM (SD=0.0426)

Results: MRMC ROC analysis

All lesions vs. normal breasts

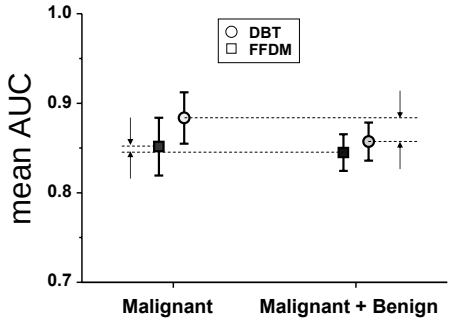
Average over 6 Readers





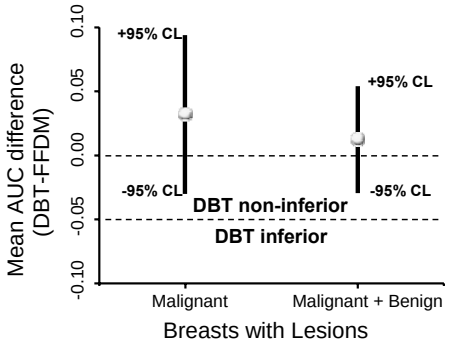
NOT SIGNIFICANTLY DIFFERENT AT 95% CL

Results: MRMC ROC analysis



DIFFERENCE IN AUCs FOR MALIGNANT VS. ALL LESIONS IS HIGHER FOR DBT THAN FOR FFDM

Results: non-inferiority



Conclusions

1. Overall clinical performance with DBT (MLO) was not significantly different vs. FFDM (CC+MLO);
2. Higher difference in AUCs for malignant vs. all lesions suggests that DBT could allow radiologists to better discriminate between malignant and benign findings.

TOMOSYNTHESIS (1-VIEW) HAS SHOWN TO BE NON-INFERIOR TO DIGITAL MAMMOGRAPHY (2-VIEWS)

Perspectives

1. SCREENING: WILL DBT REPLACE MAMMOGRAPHY?
   - Non-inferiority is insufficient (dose/cost-effectiveness)
   - Workflow needs to be proven
   - Some kind of benefit should be proven (ex. drastic reduction in recall rate – relevant in Europe ?)
2. DIAGNOSTIC: MIGHT DBT BE USEFUL AS AN ADJUNCT TO MAMMOGRAPHY?
   - Retrospective analysis on subset of data to investigate specific indications for DBT (dense breasts, architectural distortions, etc.)
   - Ensure that the same additional information cannot be easily obtained by other non-irradiating / less expensive modalities (US or 2-D extra-views).

Thank you for your attention !

gisella.gennaro@ioveneto.it