NCCPM

Optimization and performance standards of digital mammography systems

Professor David Dance

NCCPM, Royal Surrey County Hospital, Guildford, United kingdom

CANCER RESEARCH UK

CURRENT RESEARCH

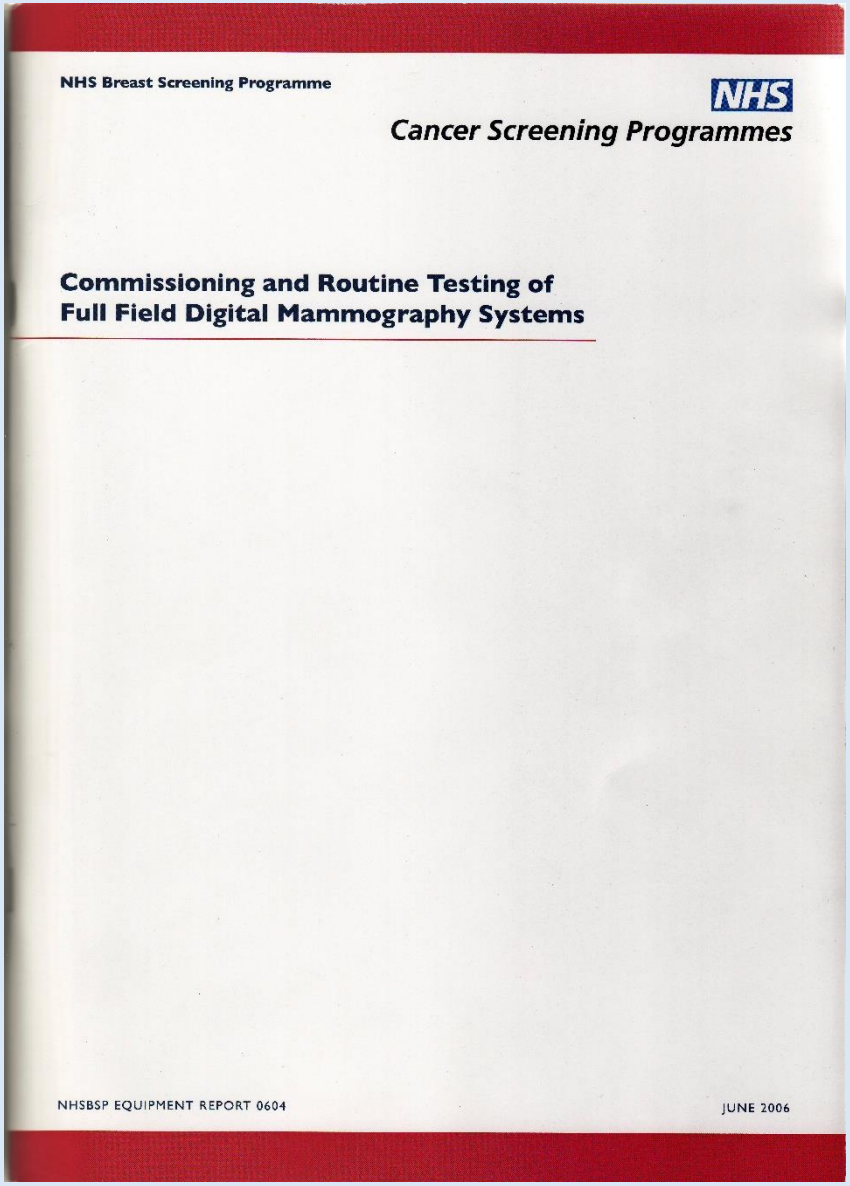
The OPTIMAM2 project:
funded by Cancer Research-UK

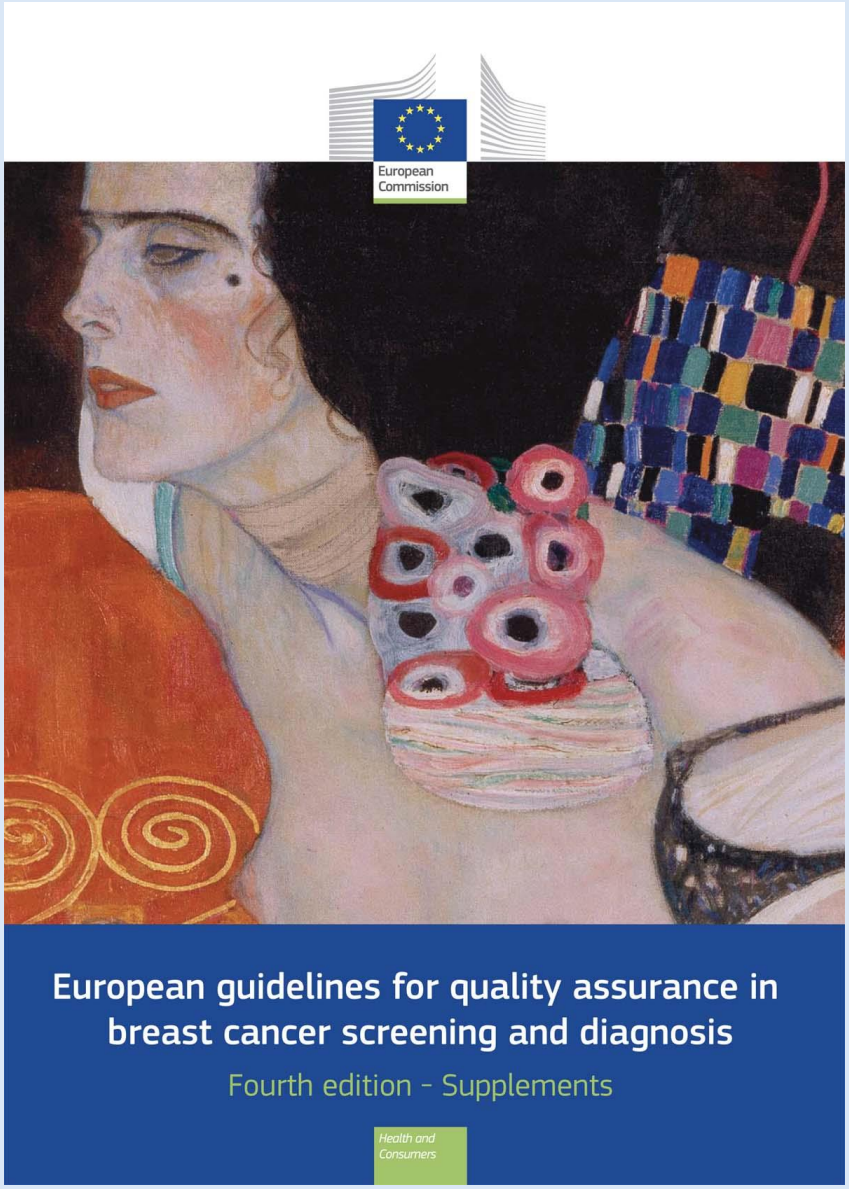
- Optimising the use of X-rays to detect breast cancers
- Investigating the performance of imaging systems using real and simulated images of breast cancers

Main Colleagues in OPTIMAM

NCCPM, Guildford	CVSSP, University of Surrey	University of Leuven	Addenbrooke's hospital, Cambridge	St George's Hospital	Jarvis Centre, Guildford
Ken Young	Kevin Wells	Hilde Bosmans	Mathew Wallis	Ros Given-Wilson	Julie Cooke
Alistair Mackenzie	Prem Elangovan	Emmy Shaheen	Paula Wilsher	Charul Patel	
Lucy Warren	Oliver Diaz	Nick Marshall			
Padraig Looney	Alaleh Rashidnasab	Lesley Cockmartin			
Mark Halling-Brown					
+ Students					

Standards and Guidelines

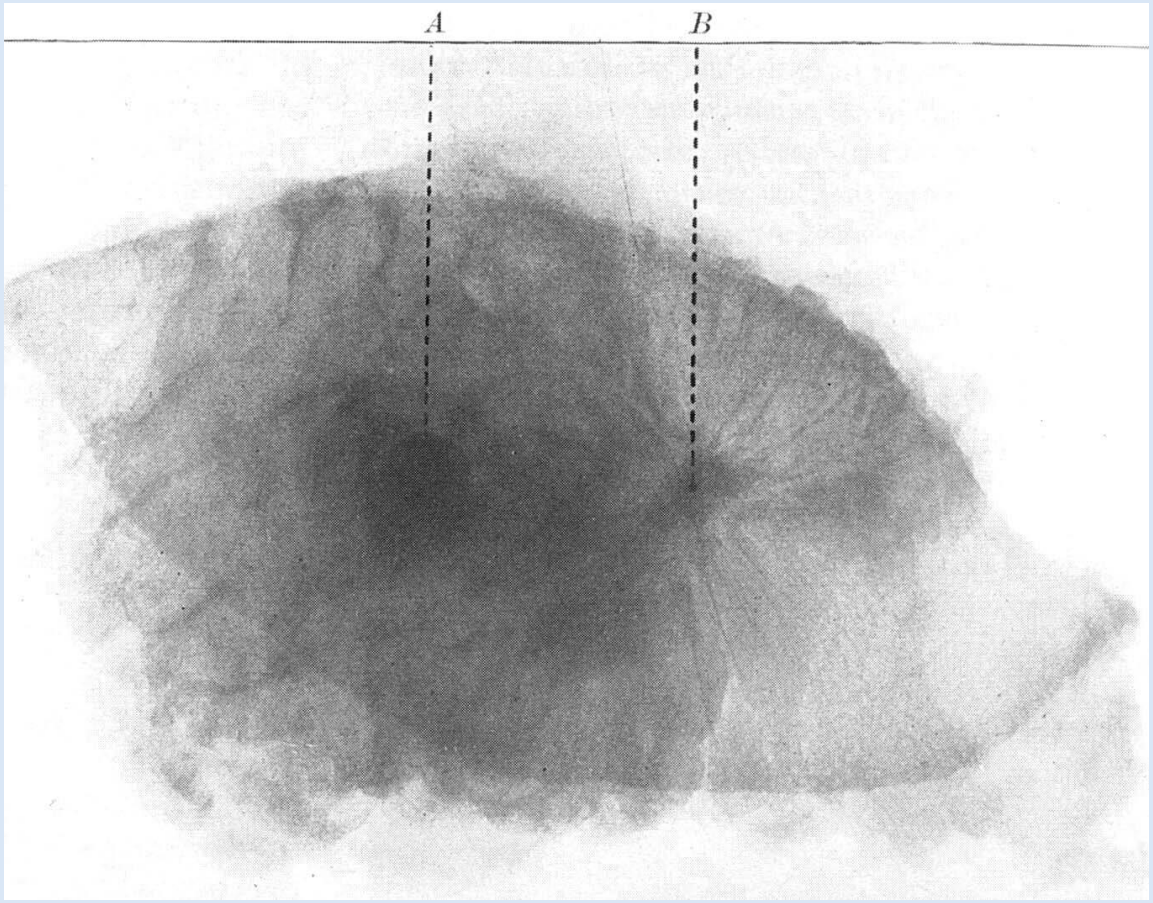




Outline

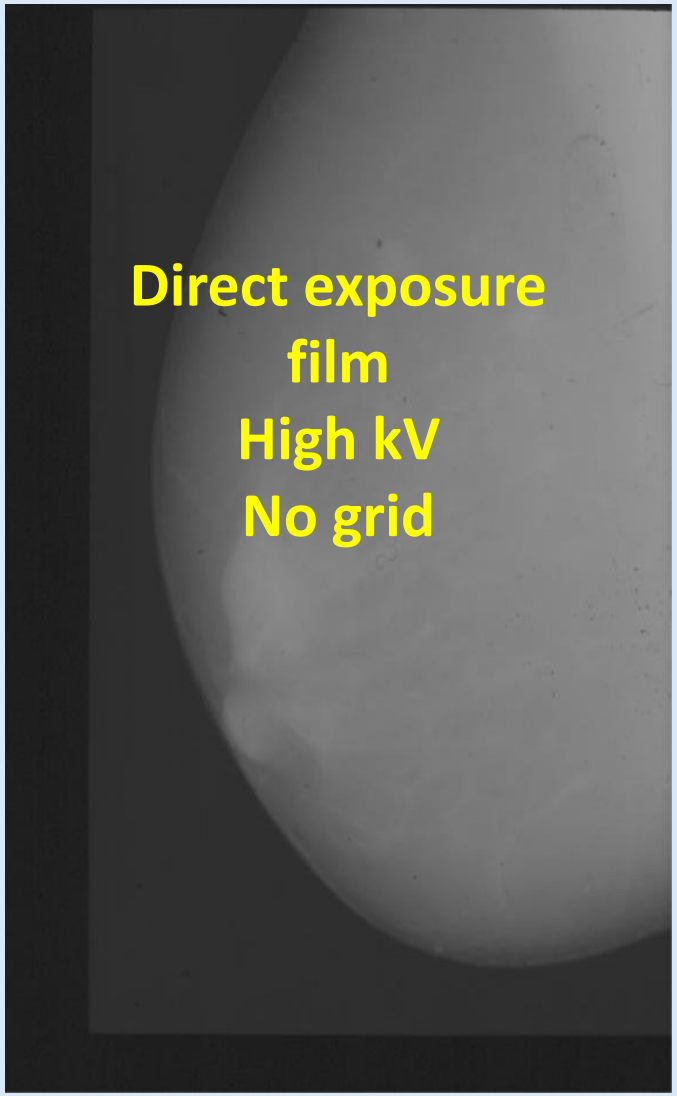
- Background
- Breast dosimetry and breast dose
- Choice of X-ray spectra
- Image quality assessment
  - Physical image quality measures
  - Quality measures using phantoms
  - Clinical measures of image quality
  - Do phantom measurements and clinical measures correlate?
- Conclusions

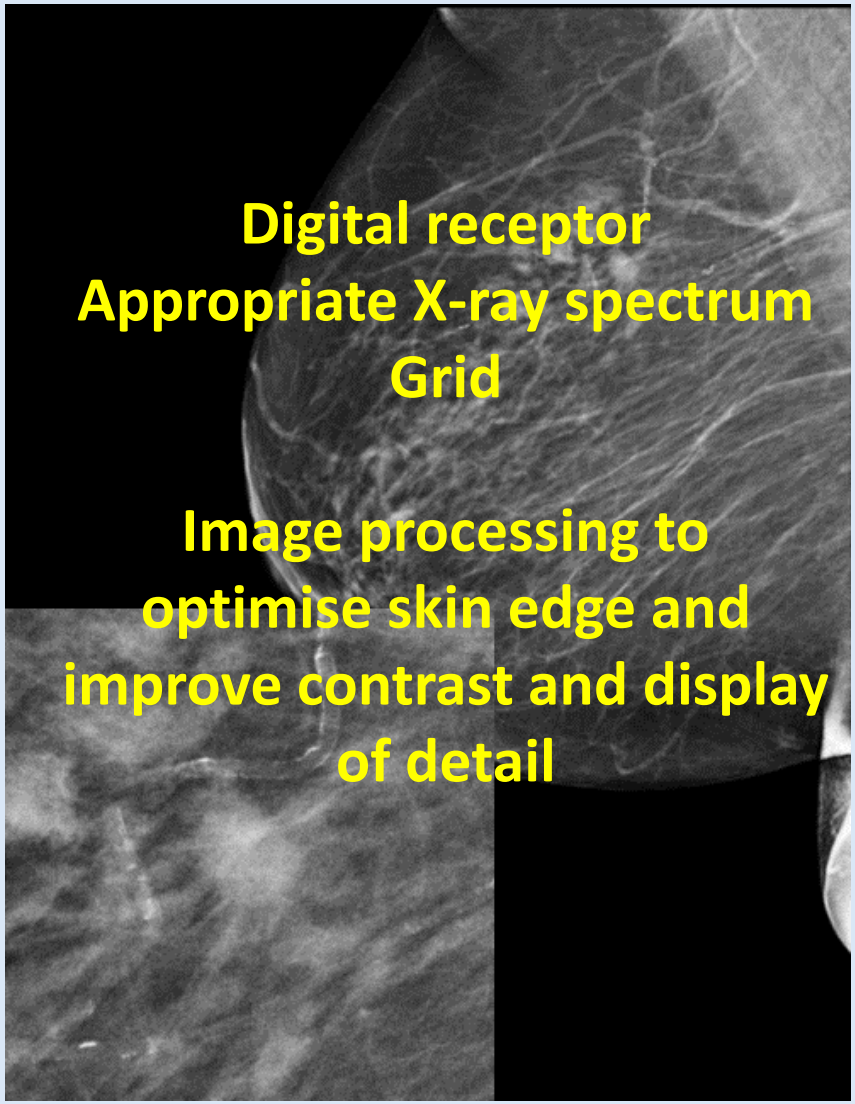
Mammography 1913



Albert Saloman (Berlin) published first radiographs of surgical breast specimens

Mammography 1980 and 2015





Mammography requirements 1

- Visualise:
  - small calcifications down to about 100 micron
    - limited by contrast, noise, system resolution & image processing
  - soft tissue masses – as small as 5mm or less
    - limited mainly by contrast, anatomical clutter (overlying tissue) and image processing

Mammography requirements 2

- Control the dose. It must be:
  - high enough so that calcs are visible against noise
  - no higher than necessary to control risk
  - BUT if it is too low, cancer detection will decrease

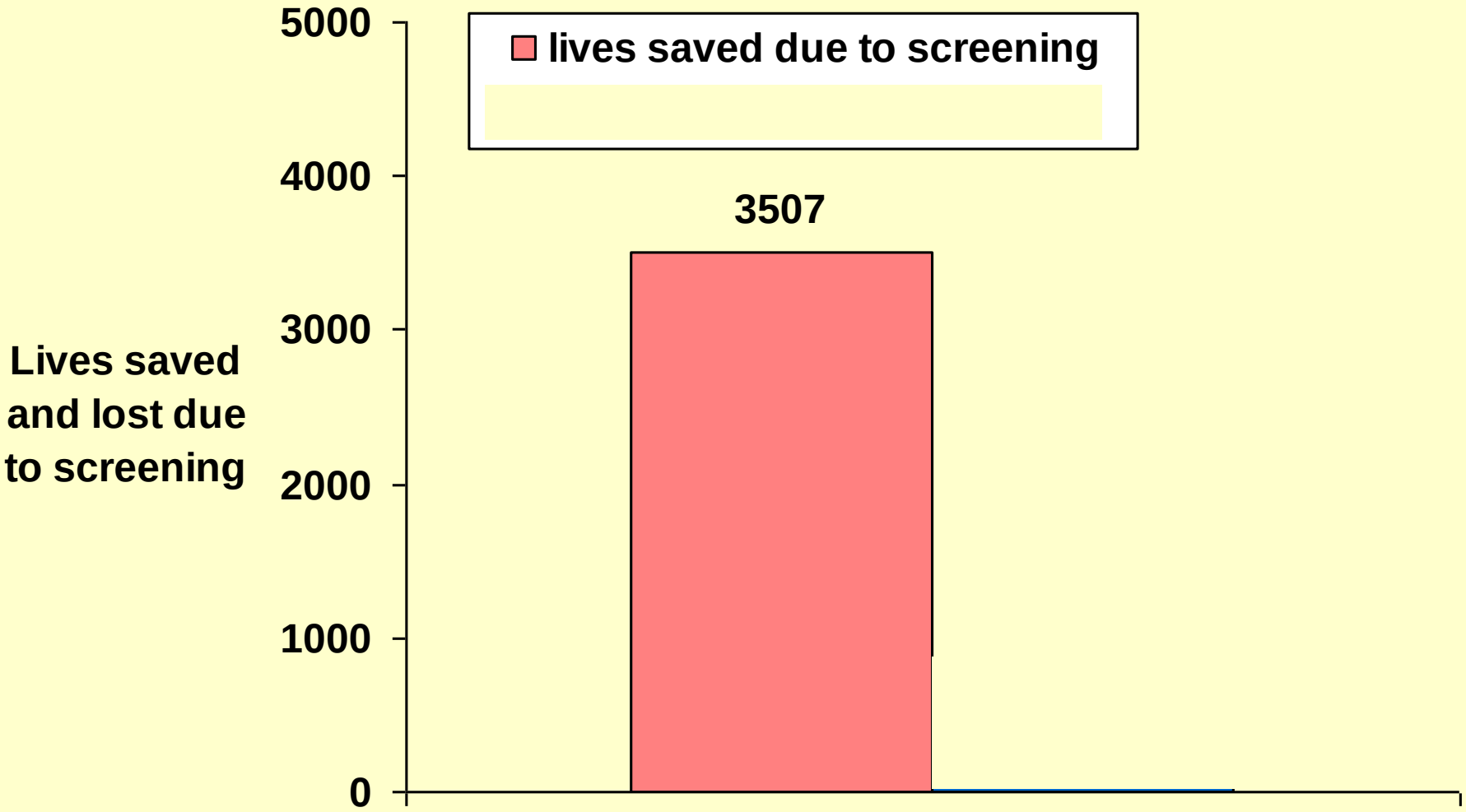
Risk Benefit Analysis

NHSBSP Report 54

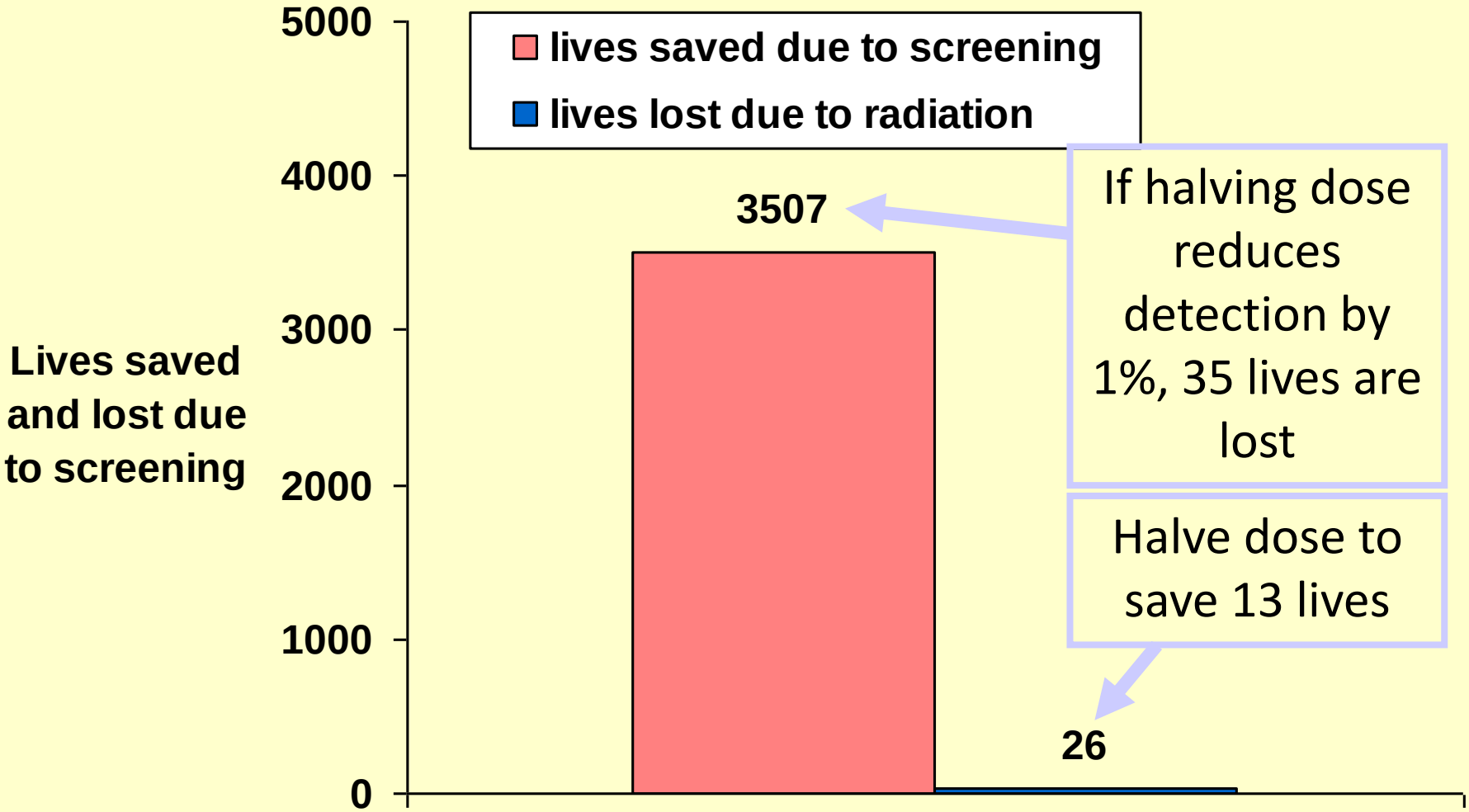
REVIEW OF RADIATION RISK IN BREAST SCREENING

Report by a joint working party of the NHSBSP National Coordinating Group for Physics Quality Assurance and the National Radiological Protection Board

Breast cancer screening
Age 50 to 70 in the UK



Breast cancer screening
Age 50 to 70 in the UK

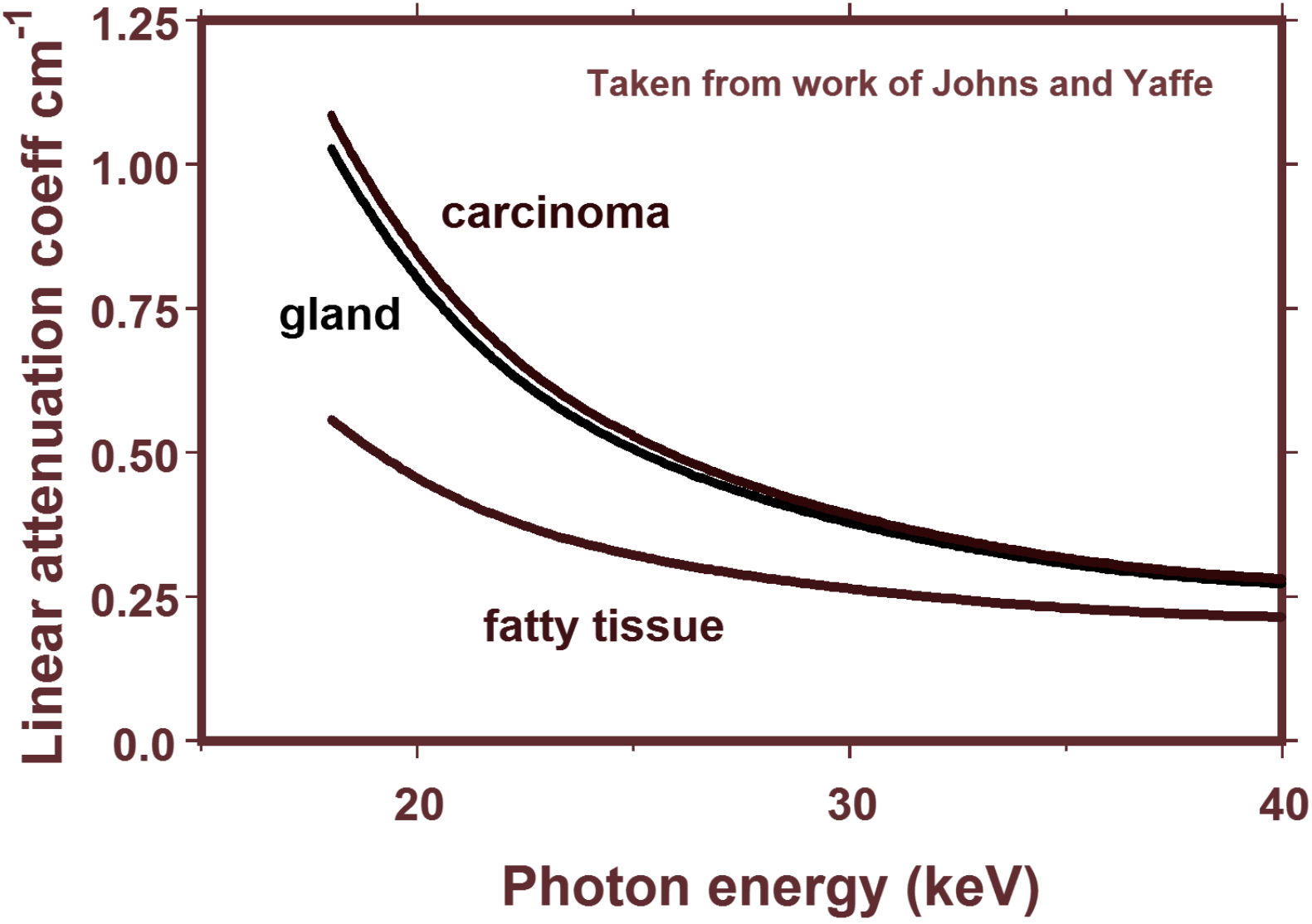


Conclusion

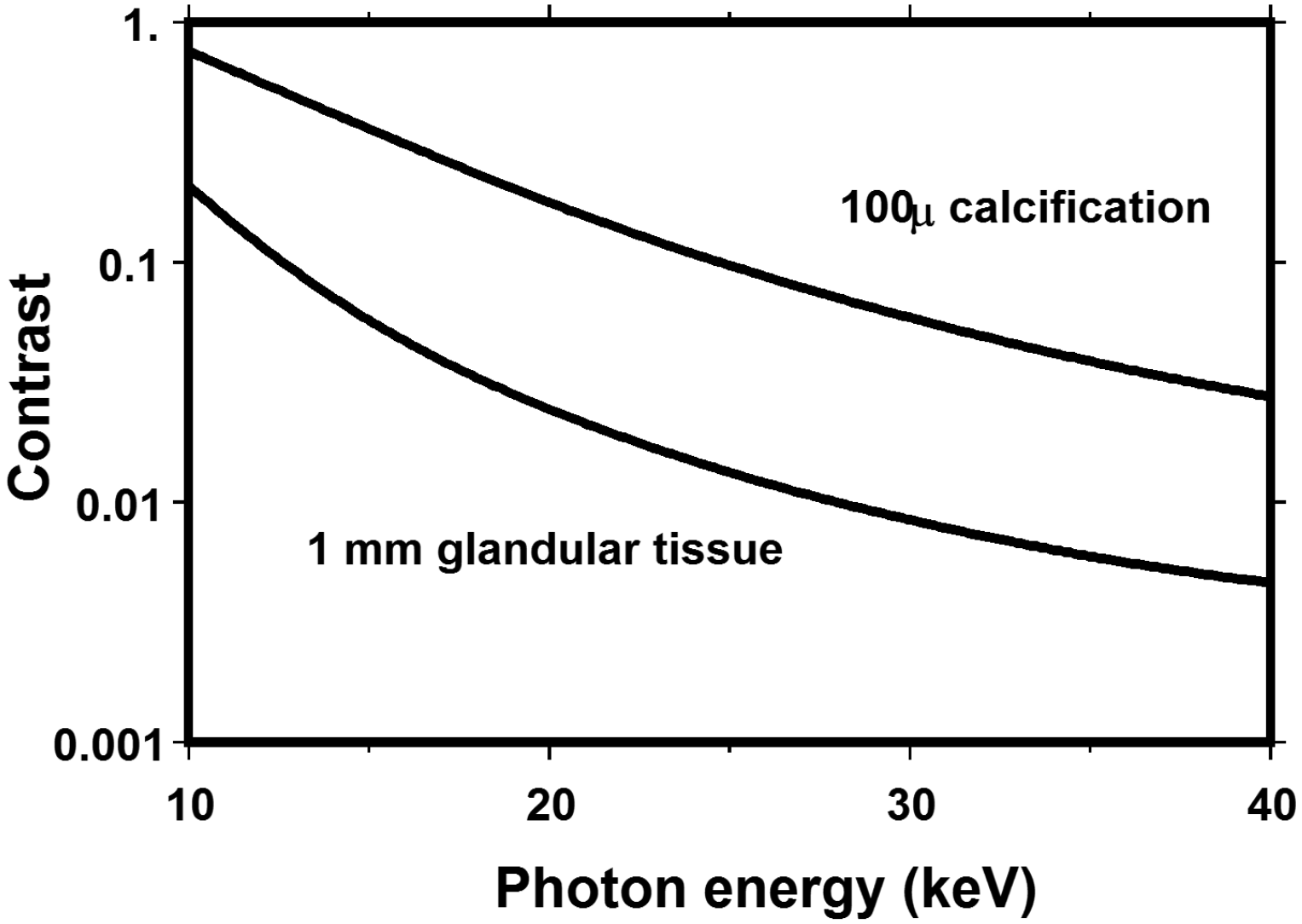
When the benefit of imaging is much greater than the radiation risk

we should concentrate on achieving sufficient image quality rather than on reducing dose

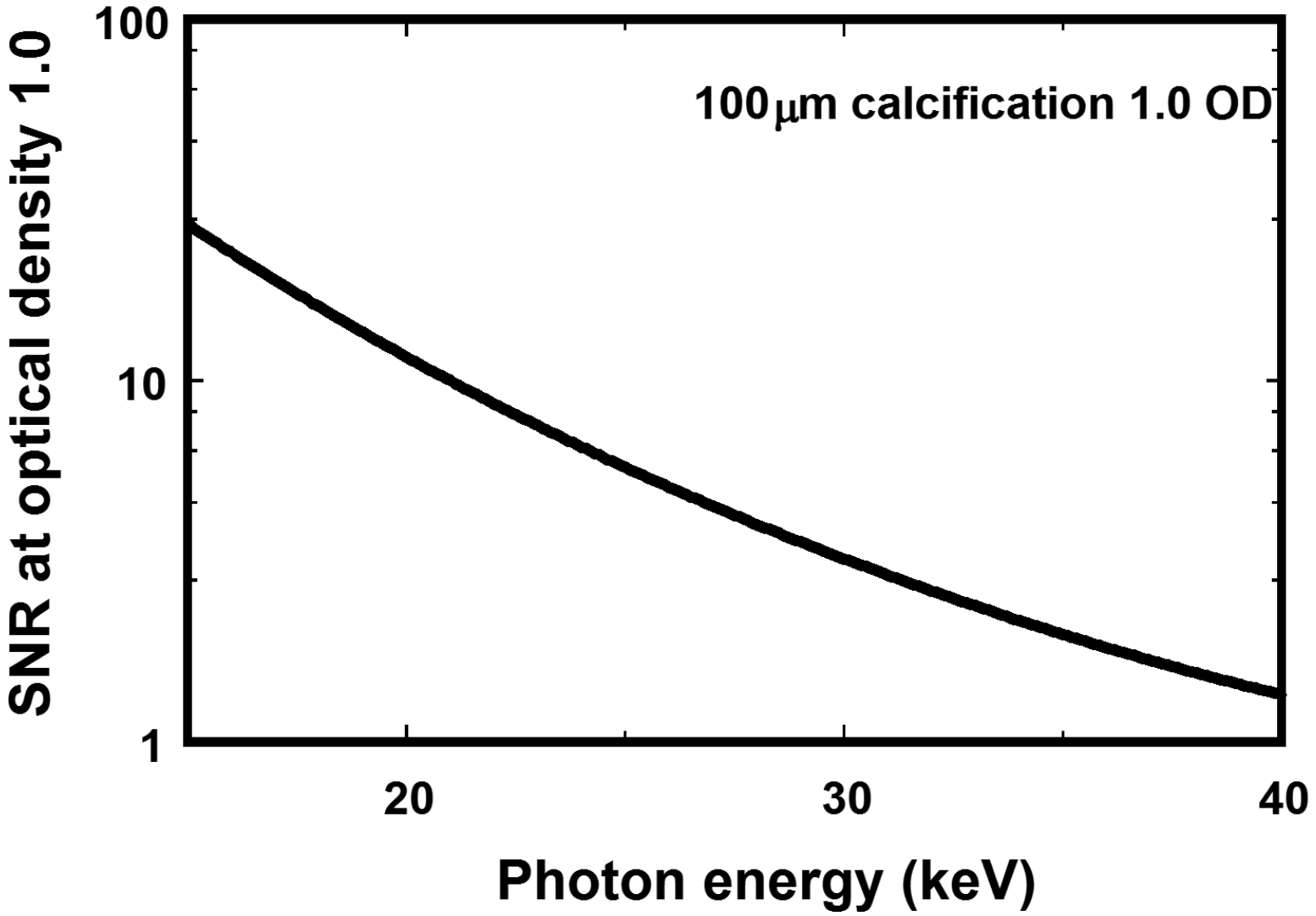
Linear attenuation coefficients



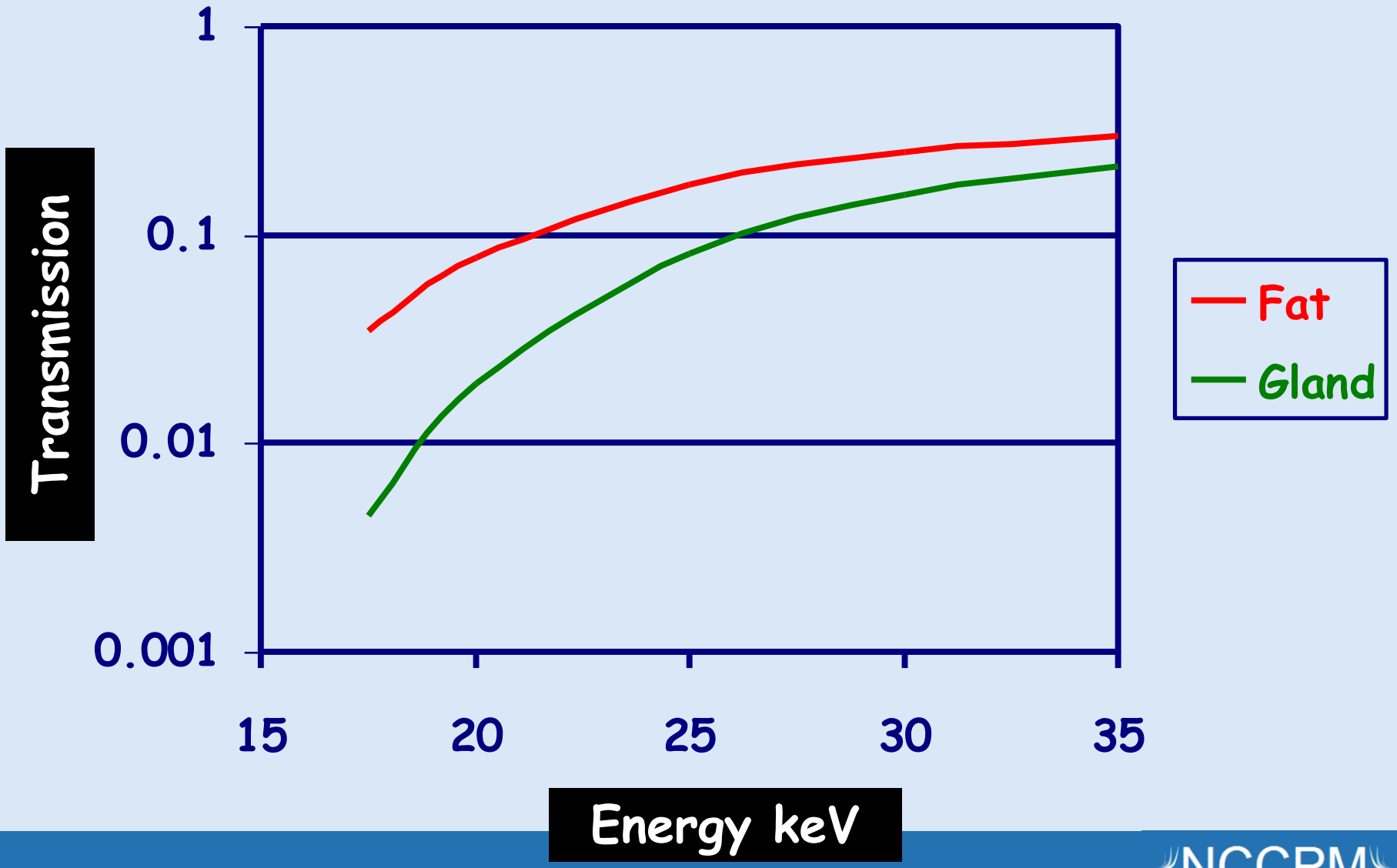
Mammographic contrast



Variation of SNR with photon energy (fixed energy deposited in receptor)



Transmission through 5 cm breast



OPTIMISATION

- A balance between dose and image quality

- Dose is influenced by
  - X-ray spectrum
  - Noise level acceptable
  - Efficiency of image receptor (DQE)

- Image quality is influenced by
  - X-ray spectrum (contrast) and dose (quantum noise)
  - Structure noise and electronic noise
  - Receptor unsharpness (MTF)
  - Focal spot size (geometric unsharpness)
  - Scatter
  - Image processing
  - Image display and viewing conditions

Outline

- Background
- Breast dosimetry and breast dose
- Choice of X-ray spectra
- Image quality assessment
  - Physical image quality measures
  - Quality measures using phantoms
  - Clinical measures of image quality
  - Do phantom measurements and clinical measures correlate?
- Conclusions

UK DOSE SURVEY 1981

- 5CM PHANTOM
- Entrance surface dose
- Dose range 0.9 to 45 mGy (!)

- kV range 24- 49 kV
- HVL range 0.2 mm Al to 1.7 mm Al

Breast dose measures

- Incident air kerma is a poor measure of breast dose

- Average glandular dose
  - Karlsson in 1976
  - Recommended by ICRP in 1987

Average glandular dose

- Cannot measure **AGD** on patients
- Incident air kerma **K** (without backscatter) can be easily measured or estimated
- Need conversion coefficients which relate **K** to AGD. In UK and Europe we use:

$$AGD = Kgcs$$

- *g*,*c* and *s* estimated using Monte Carlo modelling and simple breast model
  - tabulated against thickness and HVL
- Monte Carlo also used to design breast equivalent PMMA phantoms

D R Dance et al, PMB, 2000

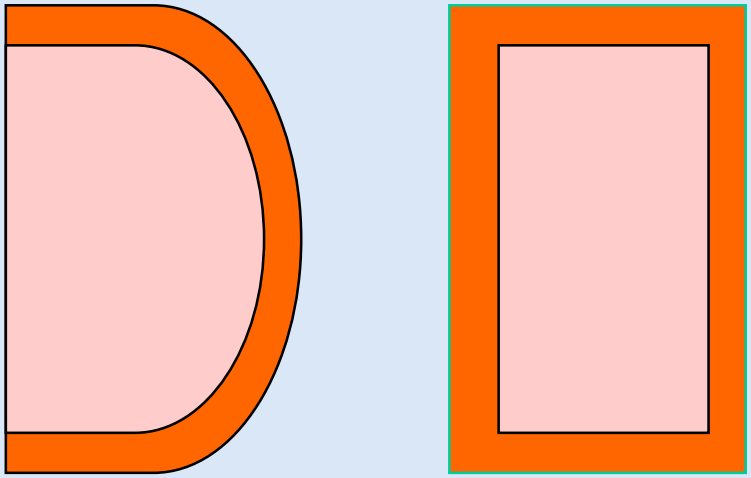
SIMPLE GEOMETRICAL BREAST MODEL

Fine for quality control of AGD using breast equivalent phantoms or series of patient data

Does not give the true dose for individual patients

Simple breast model

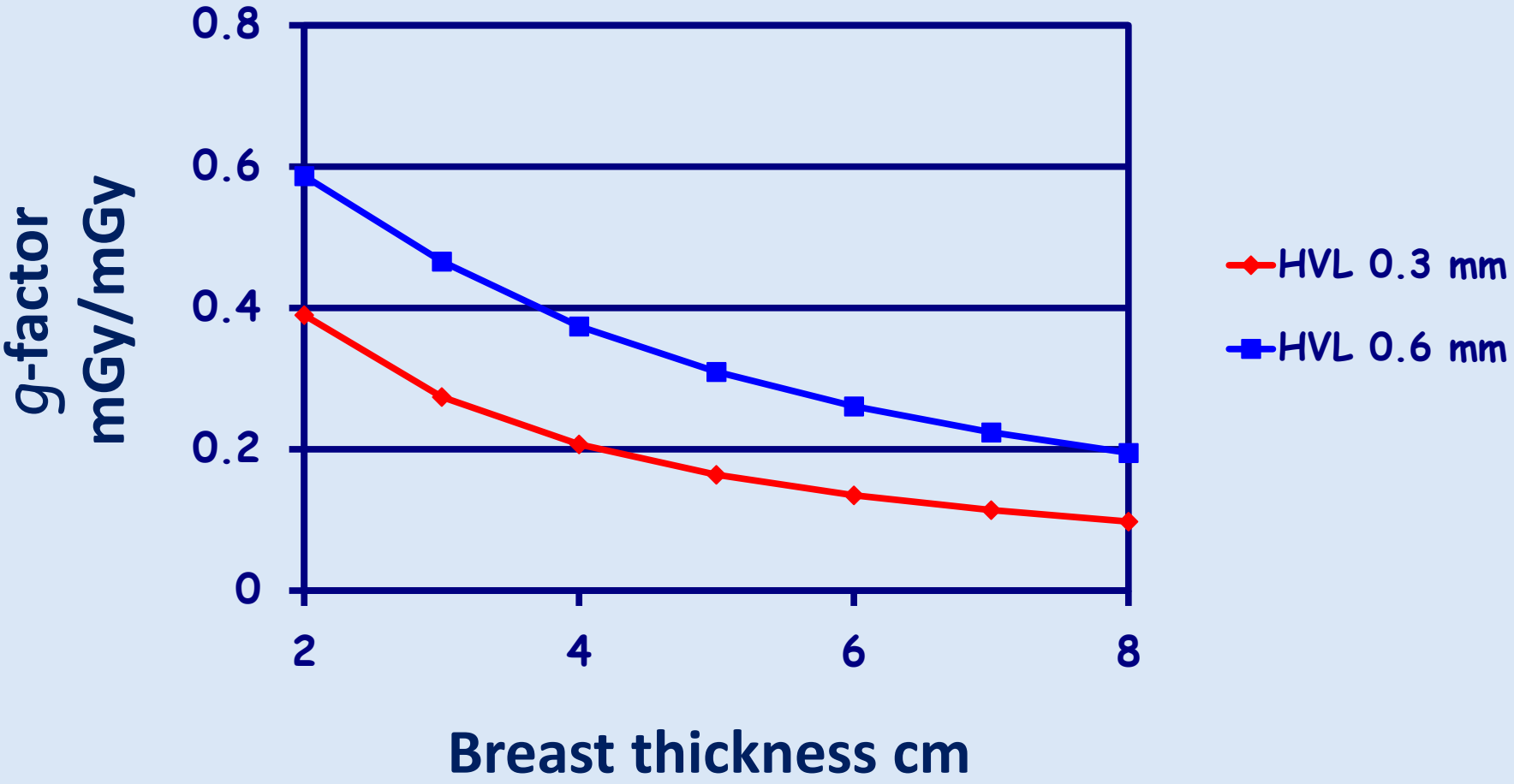
Hammerstein 1979



- 5 mm adipose shield region
- Central region with mixture of glandular and adipose tissues
- Fraction by weight of glandular tissue in central region is known as the glandularity

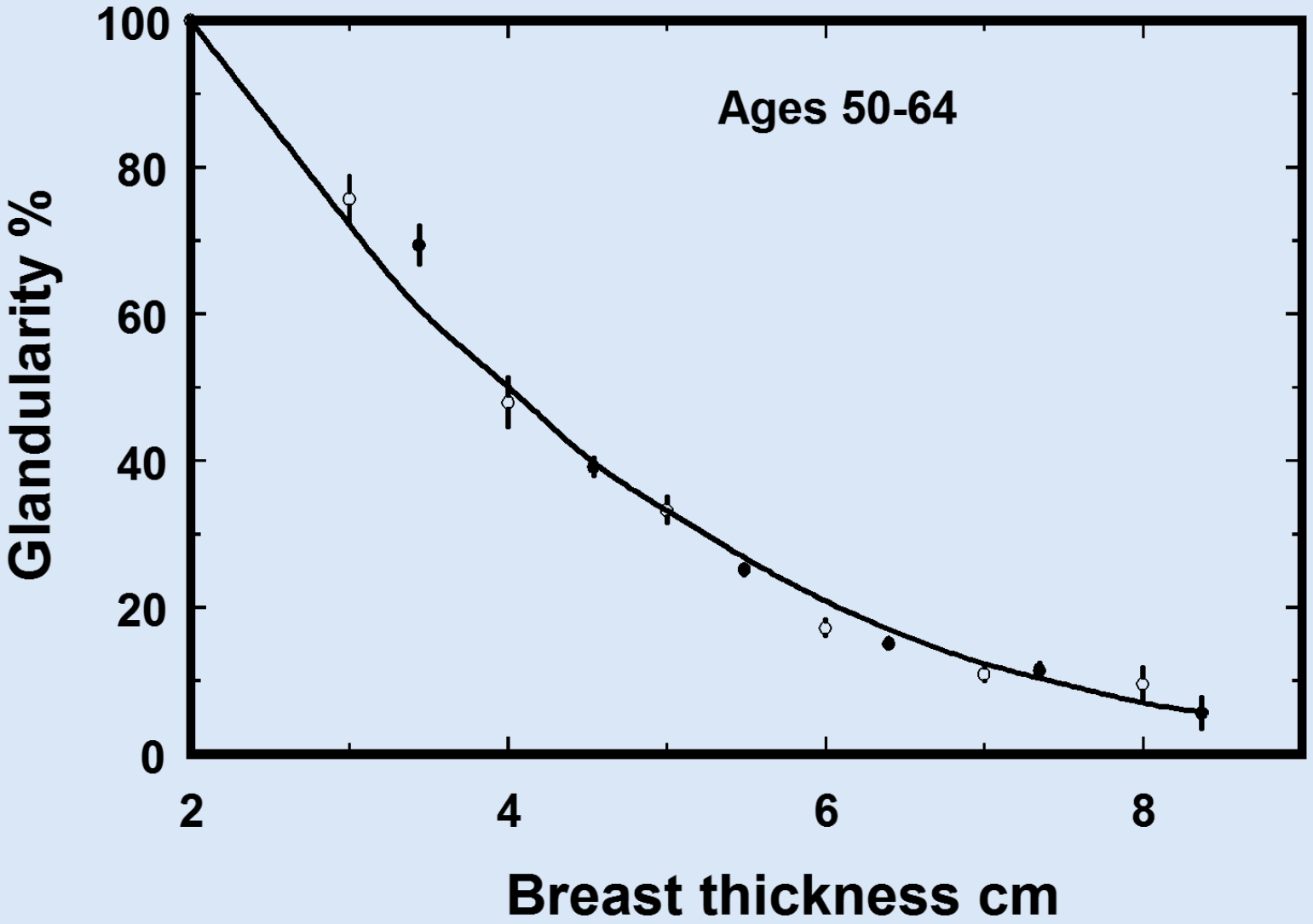
Hammerstein et al suggested the use of 50% glandularity for dosimetry

g-factors



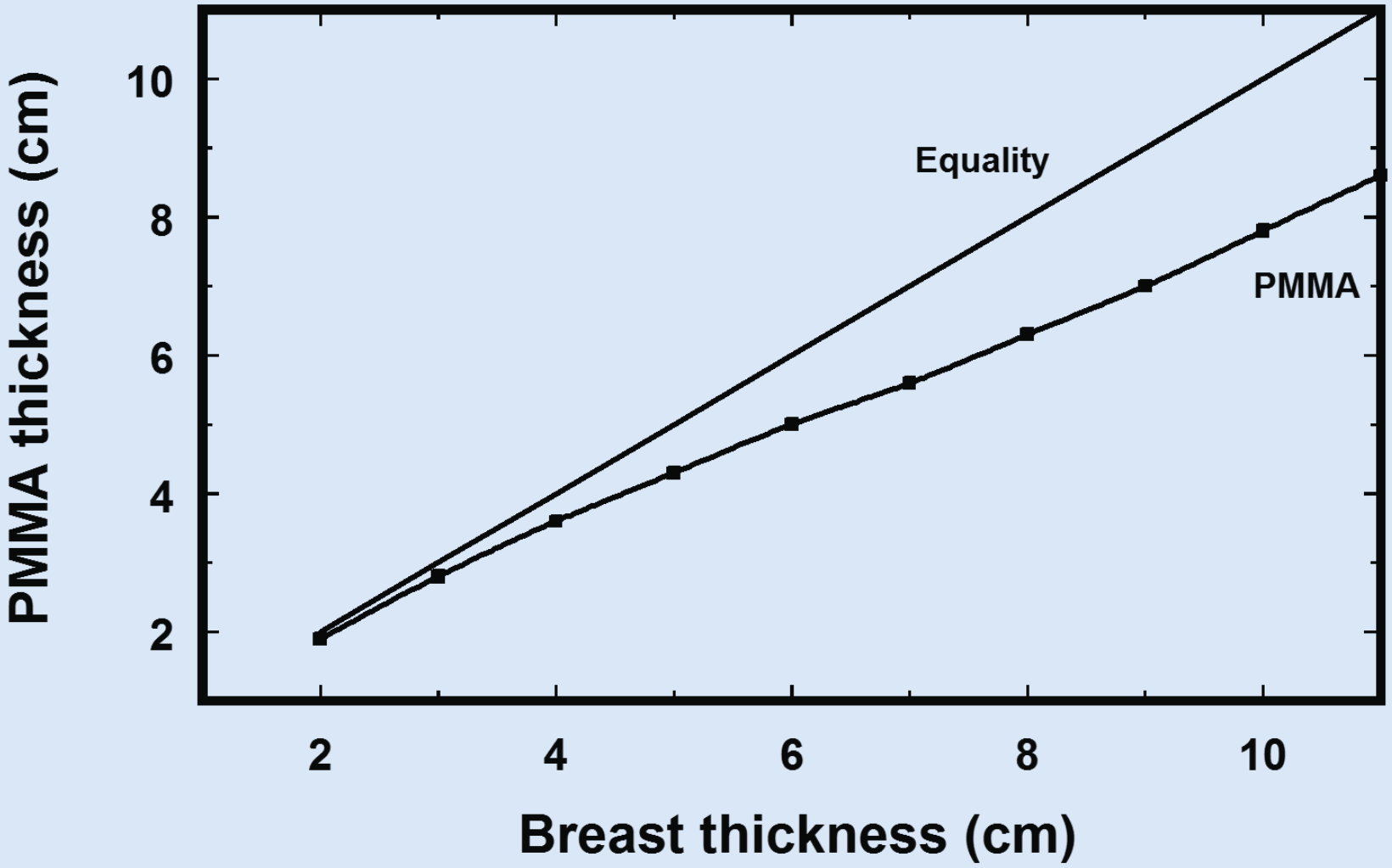
D R Dance, PMB, 1990

Breast glandularity (UK)



D R Dance et al, PMB, 2000

PMMA equivalence age 50-64



D R Dance et al, PMB, 2000

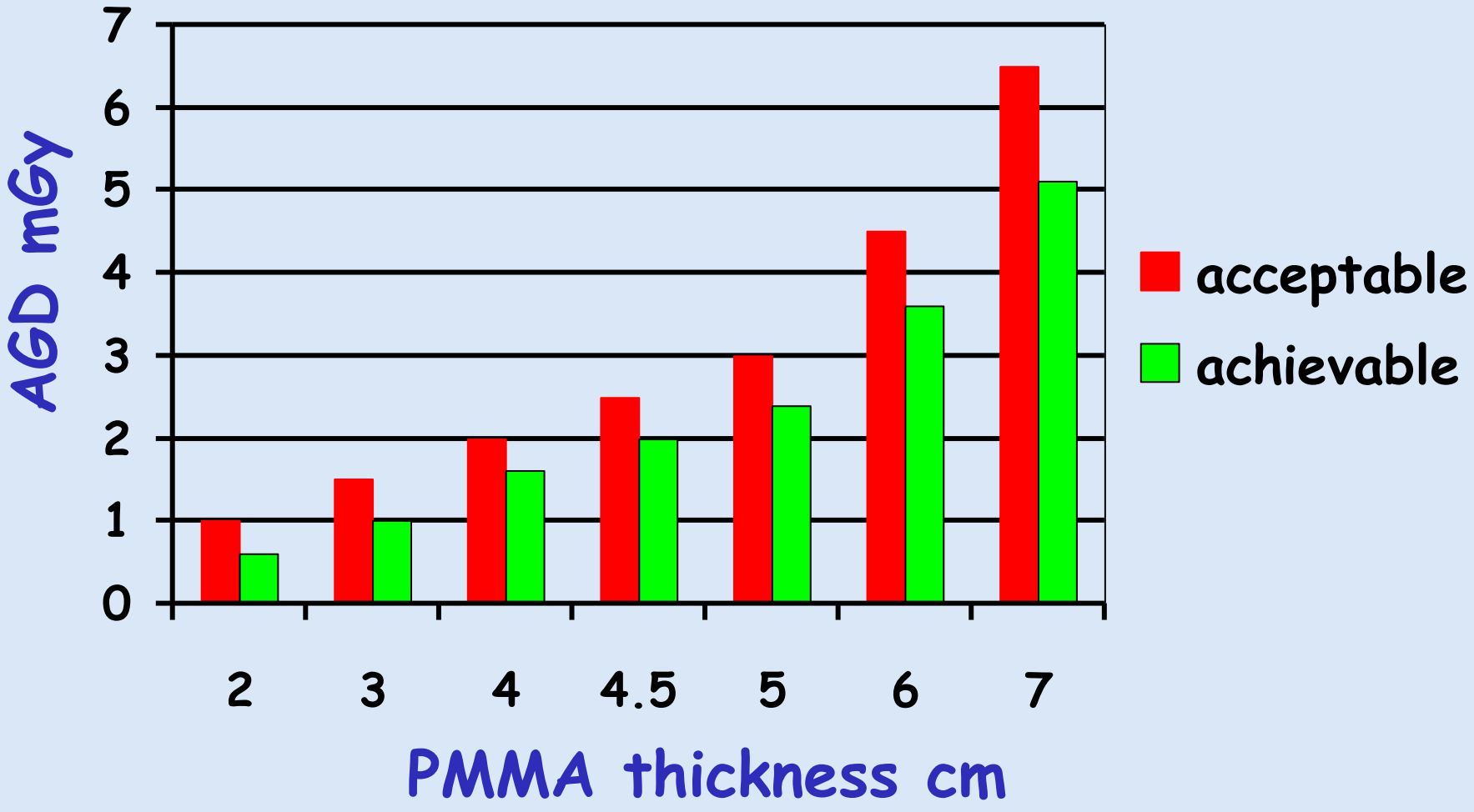
EUREF dose standards for digital mammography

- Use PMMA phantoms 20-70 mm thick
- Clinically selected AEC and spectra
- Based on previous standards for screen-film and survey data from Holland, UK and Germany

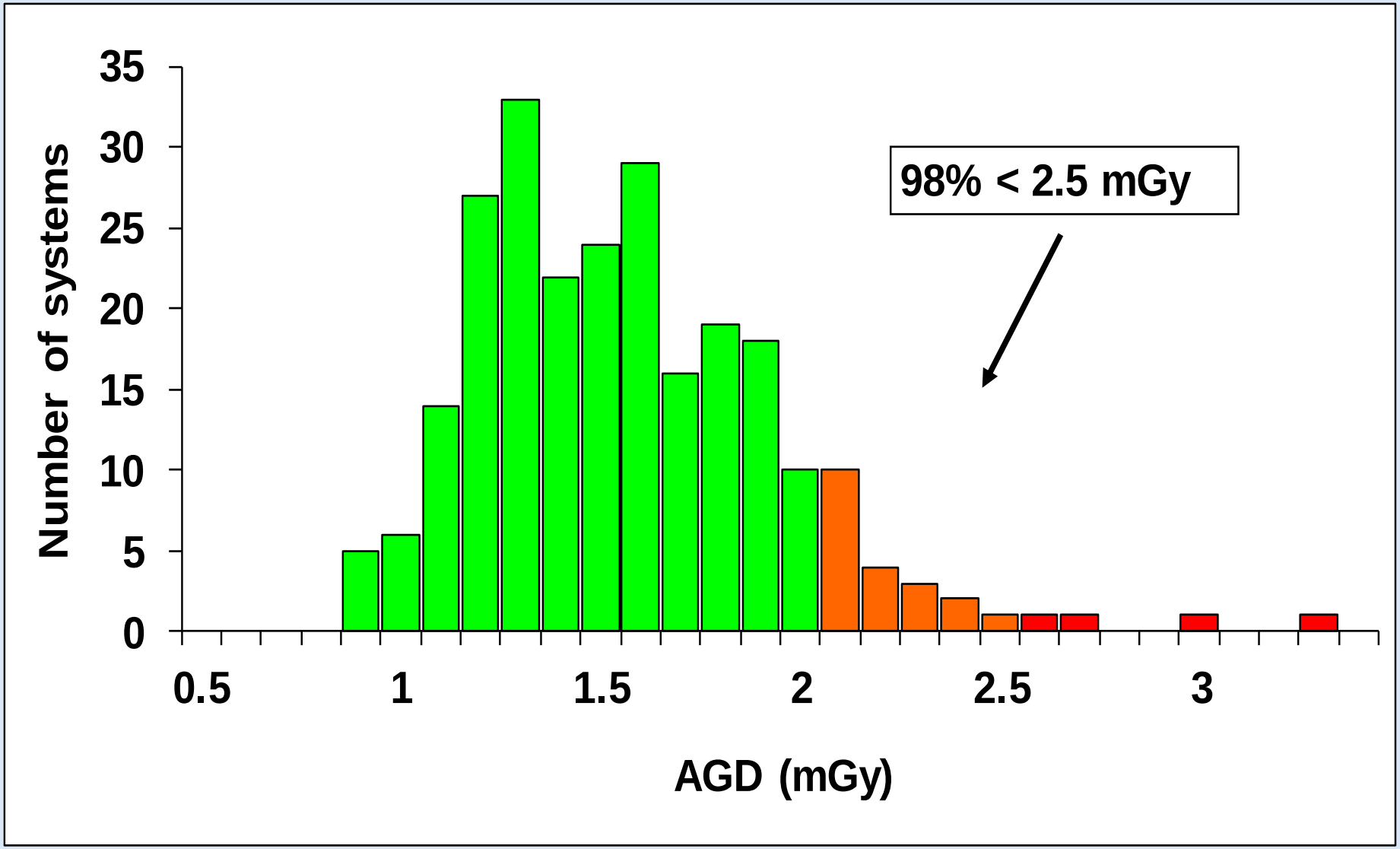
- The acceptable level is the minimum acceptable standard
- However, it is recommended that systems operate as far as possible at a standard equal to or better than the achievable level

- REMEMBER that if the dose is too low, cancer detection will decrease
  - different receptors have different unsharpness and noise properties and will require different doses

Acceptable and achievable AGD levels



AGD: 53 mm breast (45 mm PMMA)

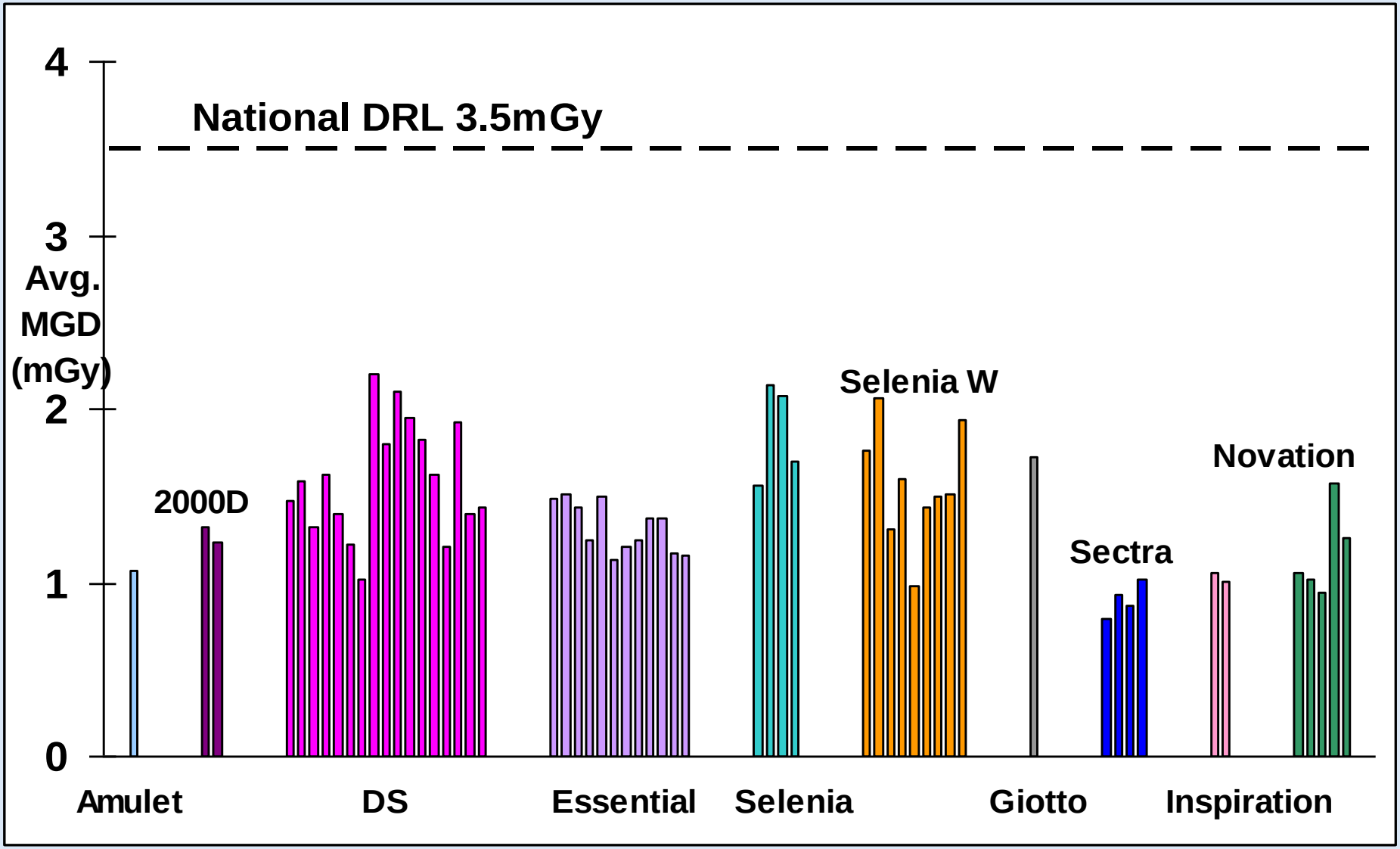


Patient dose

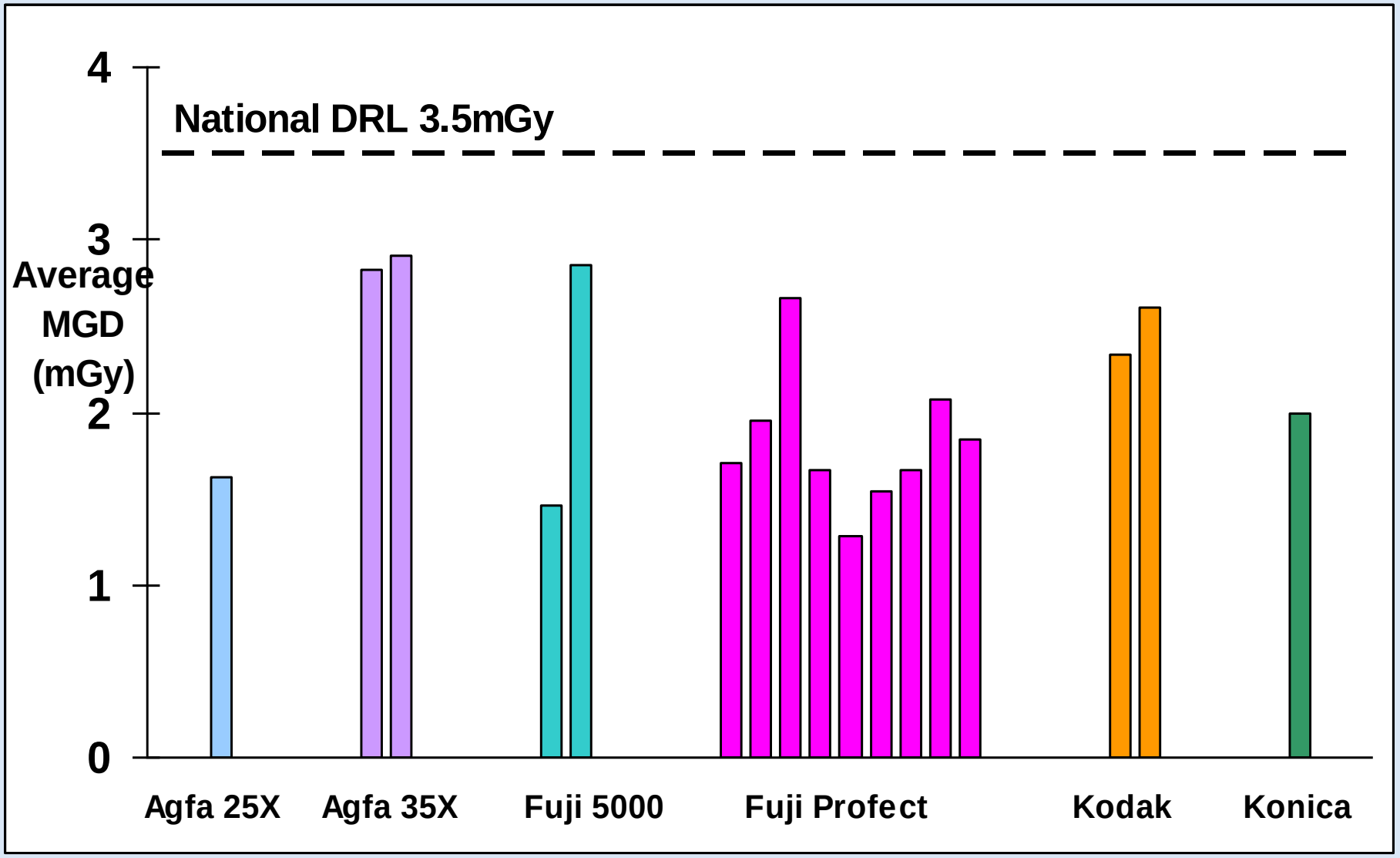
- **What dose level is currently used?**

Reference: Oduko J et al "A survey of patient doses from digital mammography systems in the UK in 2007 to 2009" in Proceedings of the 10th International Workshop on Digital Mammography 2010

DR systems, 50-60mm, OB views



CR systems, 50-60mm, OB views



Patient dose

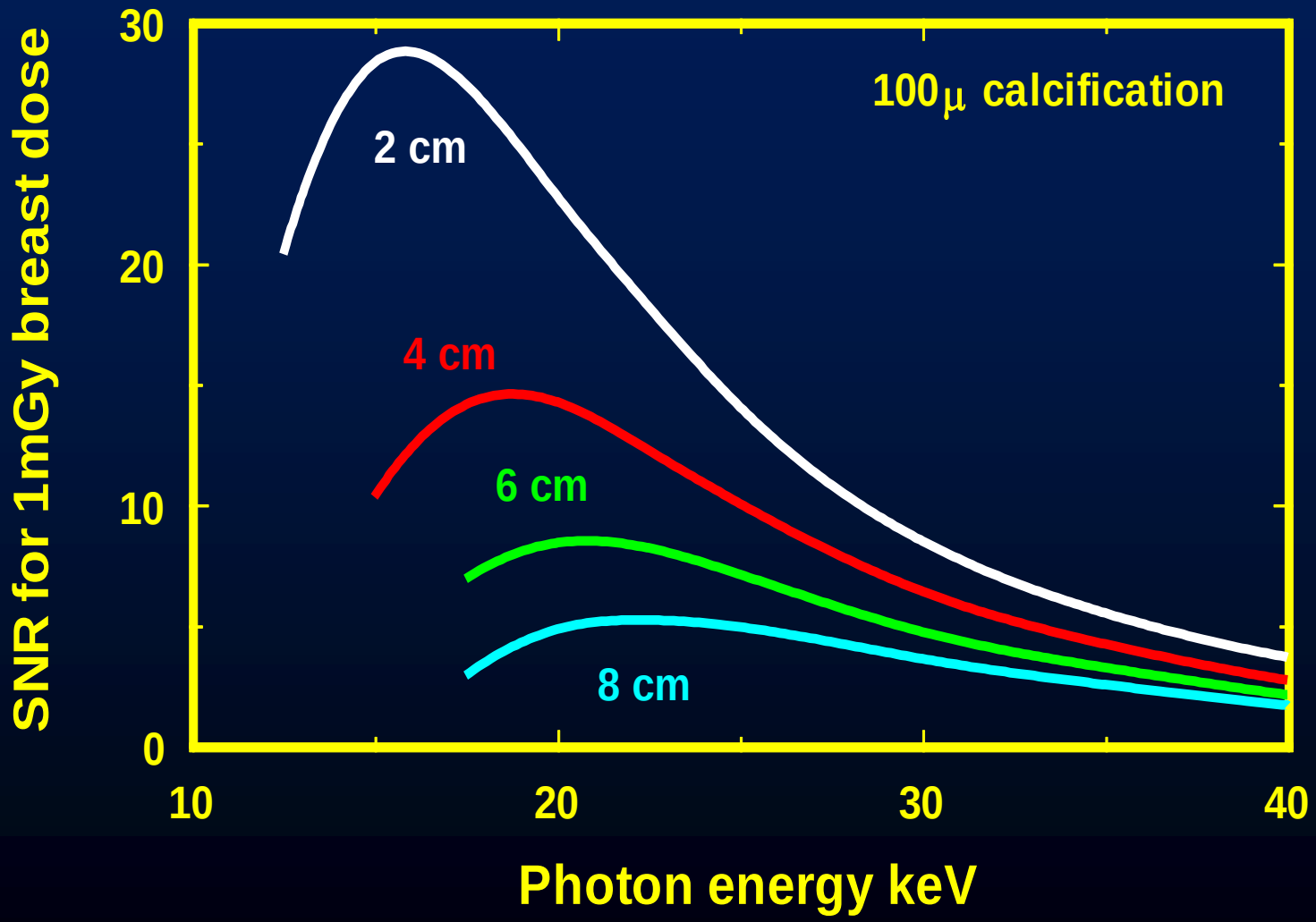
- What dose level is currently used?
  Answer: 0.8 – 2.8 mGy AGD for 50-60mm breasts
- Doses for CR are generally higher than for DR
- Different systems have differ noise and unsharpness and require different dose levels
- What dose level should be used?

We can measure how changing the dose changes the results of physical or test phantom measurements BUT how does it affect cancer detection?

Outline

- Background
- Breast dosimetry and breast dose
- Choice of X-ray spectra
- Image quality assessment
  - Physical image quality measures
  - Quality measures using phantoms
  - Clinical measures of image quality
  - Do phantom measurements and clinical measures correlate?
- Conclusions

Optimal energies for mammography



Optimal energy (simple model)

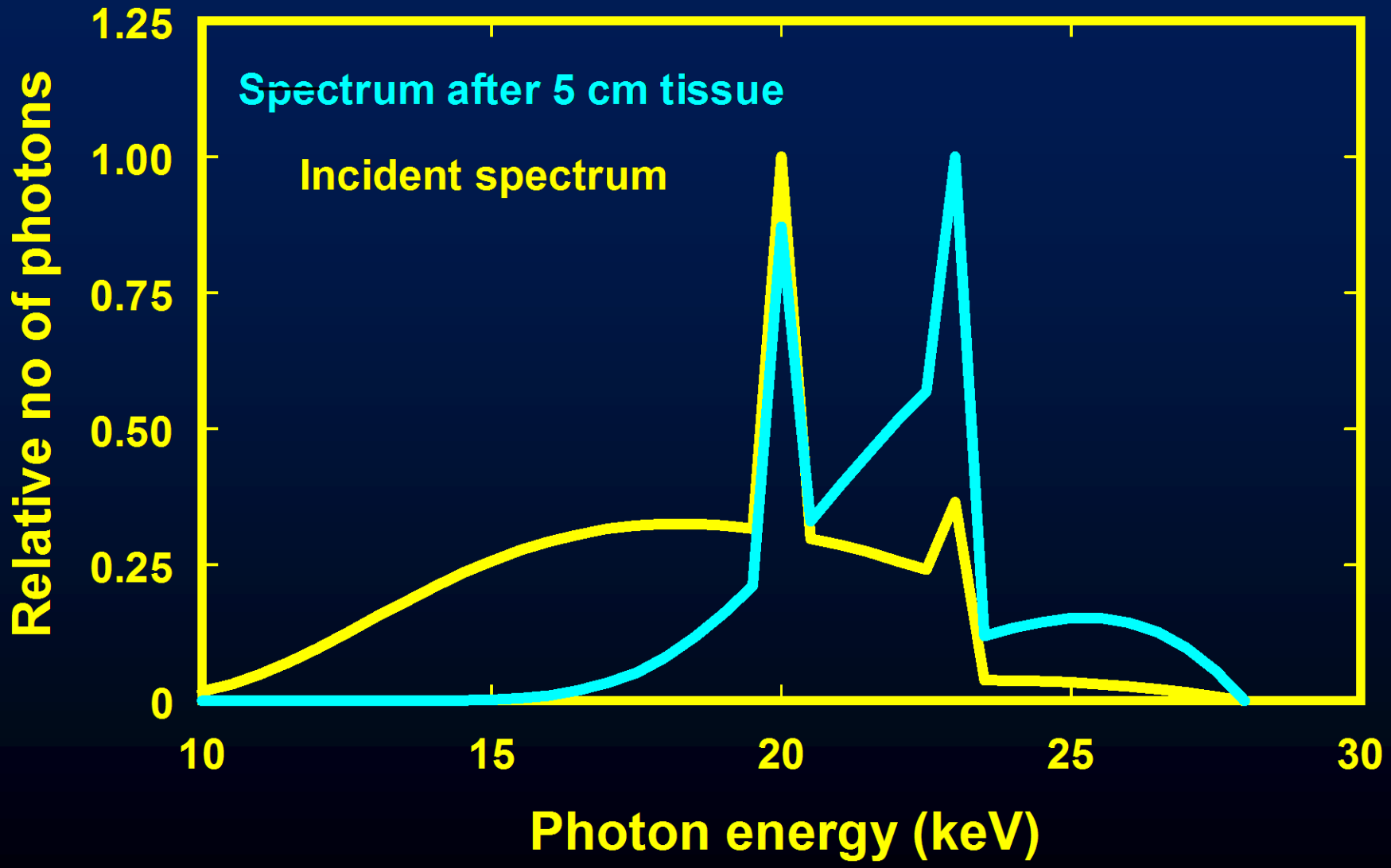
Varies between approximately 16 - 23 keV

Increases with breast thickness

How can we adjust the spectrum?

- Mo target:
  - K X-rays 17.4 and 19.6 keV
- Rh target:
  - K X-rays 20.2 and 22.7 keV
- W target:
  - K X-rays 59.3 and 67.2 keV

Rh/Rh spectra

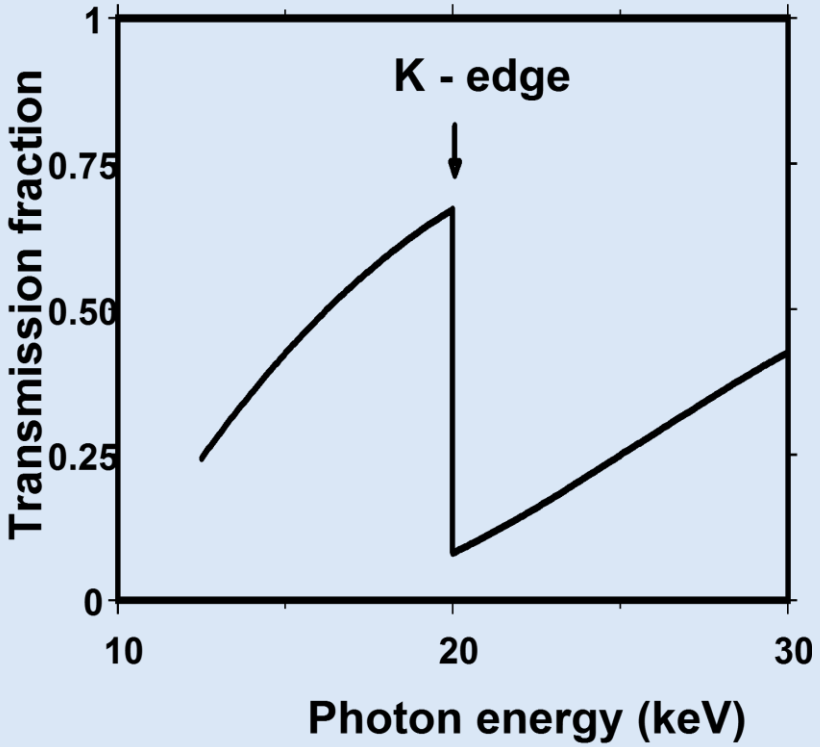


K-edges

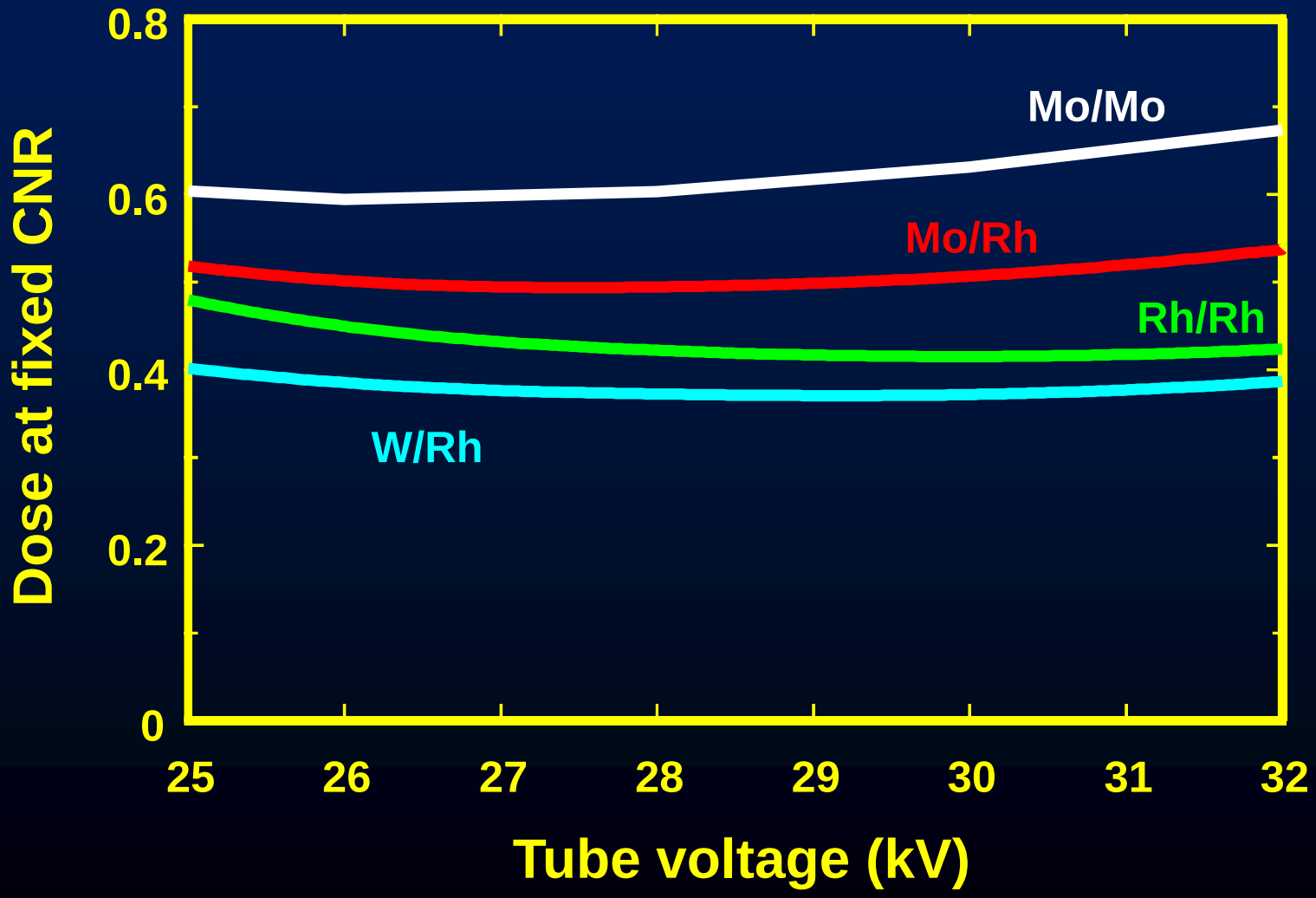
Mo 20.0 keV

Rh 23.3 keV

Ag 25.5 keV



8 cm breast 10% glandularity



Spectra for digital mammography

- Optimum spectrum varies with breast thickness and glandularity
- Mo/Mo is optimum only for 2 cm breasts
- Dose saving possible cf Mo/Mo when CNR (not contrast) is main constraint

Spectra for digital receptors:

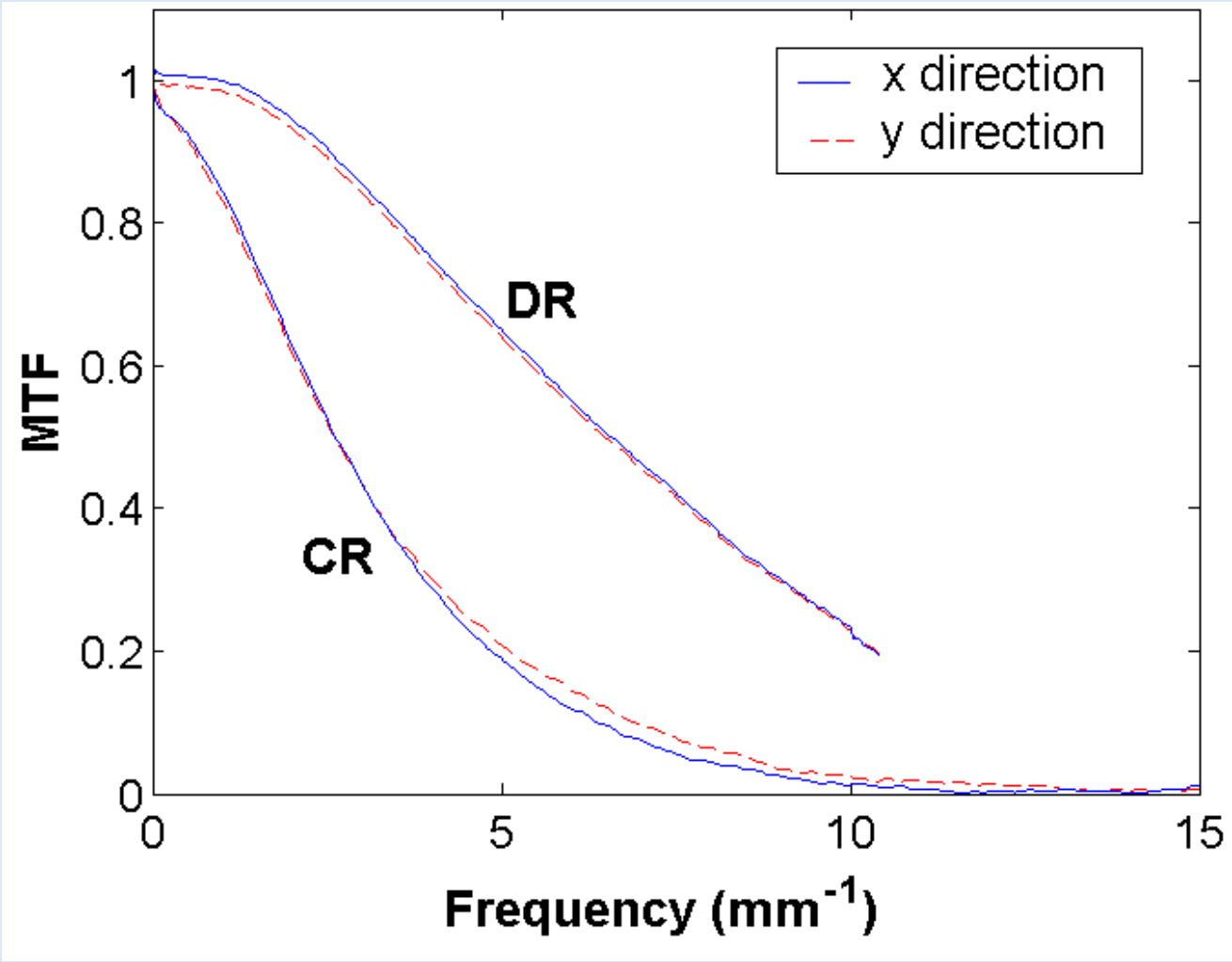
For the same generating kV:

Anode/Filter	Beam Energy	Breast size
Mo/Mo	Lowest	Smallest
Mo/Rh		
Rh/Rh		
W/Rh		
W/Ag	Highest	Largest

Outline

- Background
- Breast dosimetry and breast dose
- Choice of X-ray spectra
- Image quality assessment
  - Physical image quality measures
  - Quality measures using phantoms
  - Clinical measures of image quality
  - Do phantom measurements and clinical measures correlate?
- Conclusions

Modulation transfer functions



MTF for DR is better than for CR

Noise in digital images

Arises from:

Quantum mottle

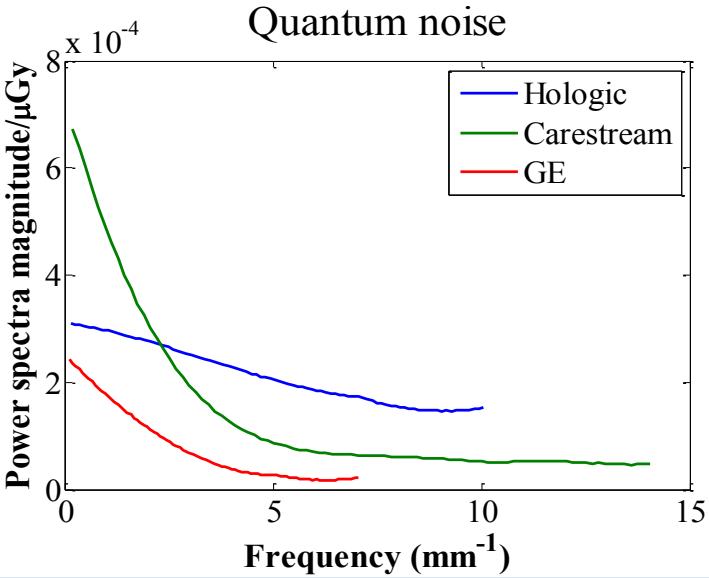
- fluctuations in no of X-ray photons detected and no of light photons/charge carriers produced per X-ray photon

Structure mottle

- generally small

Electronic sources

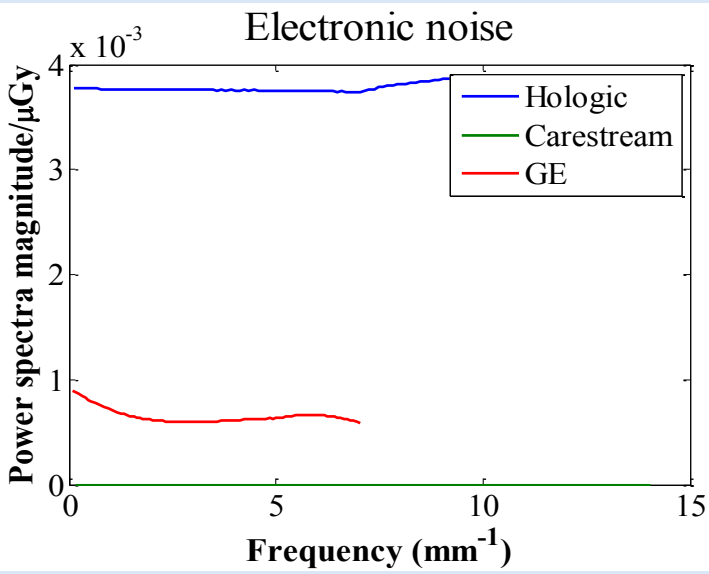
Relative noise power spectra contributions

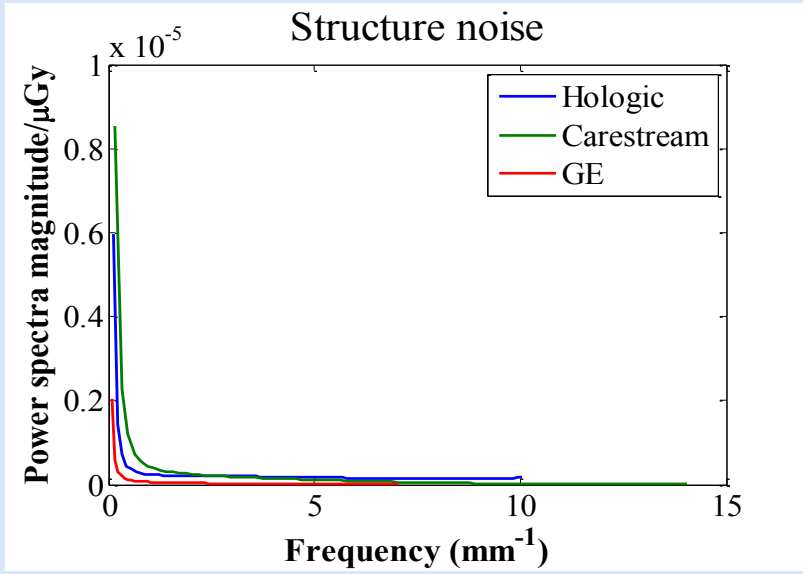


Quantum is generally largest overall contribution

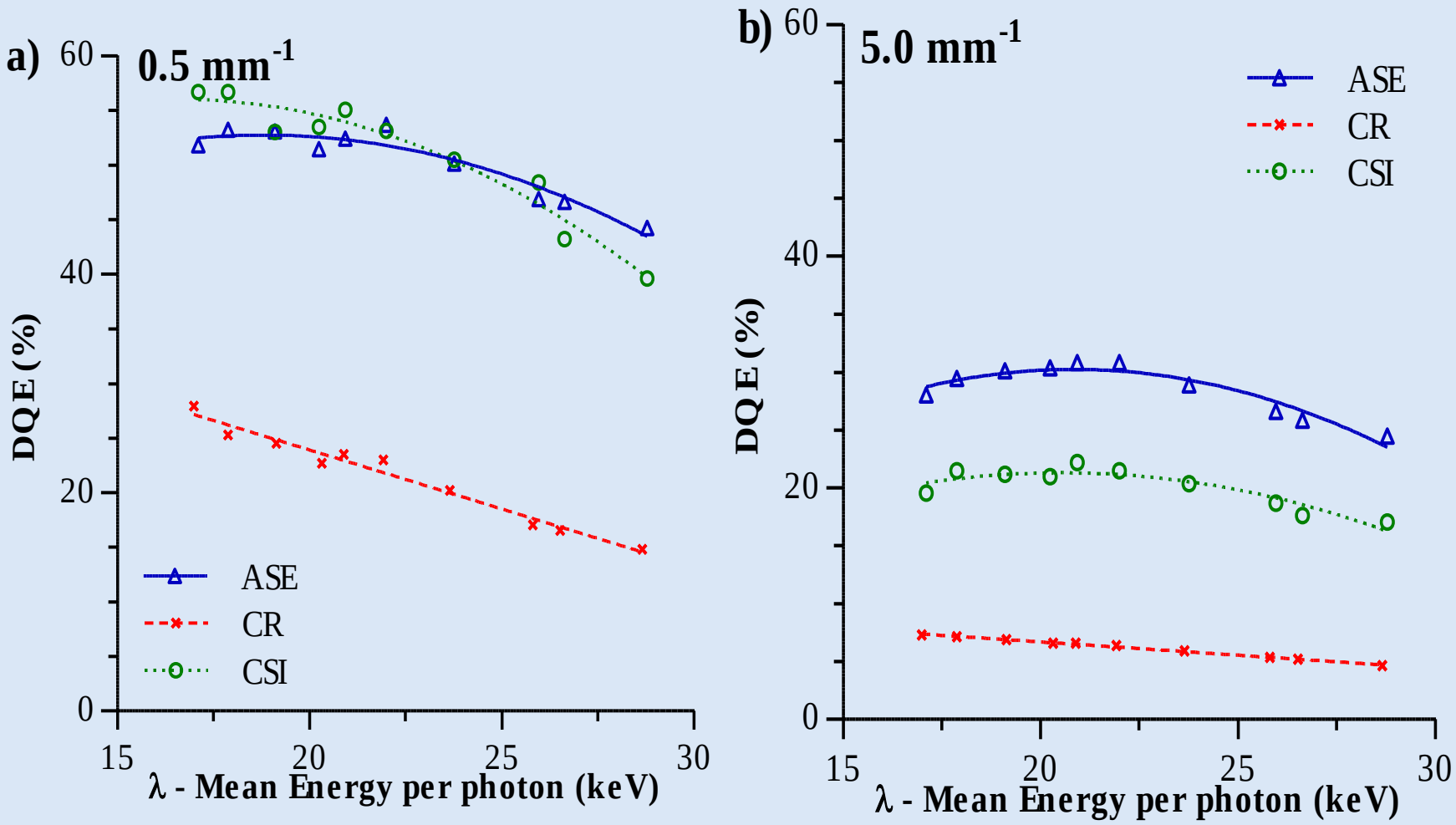
Low frequency quantum is largest for CR

Quantum for amorphous Se decreases less with frequency





DQE comparison



DQE for DR better than for CR especially at higher frequency

How does the difference between MTF and DQE for CR and DR impact on:

Image quality assessed using test phantoms

Outline

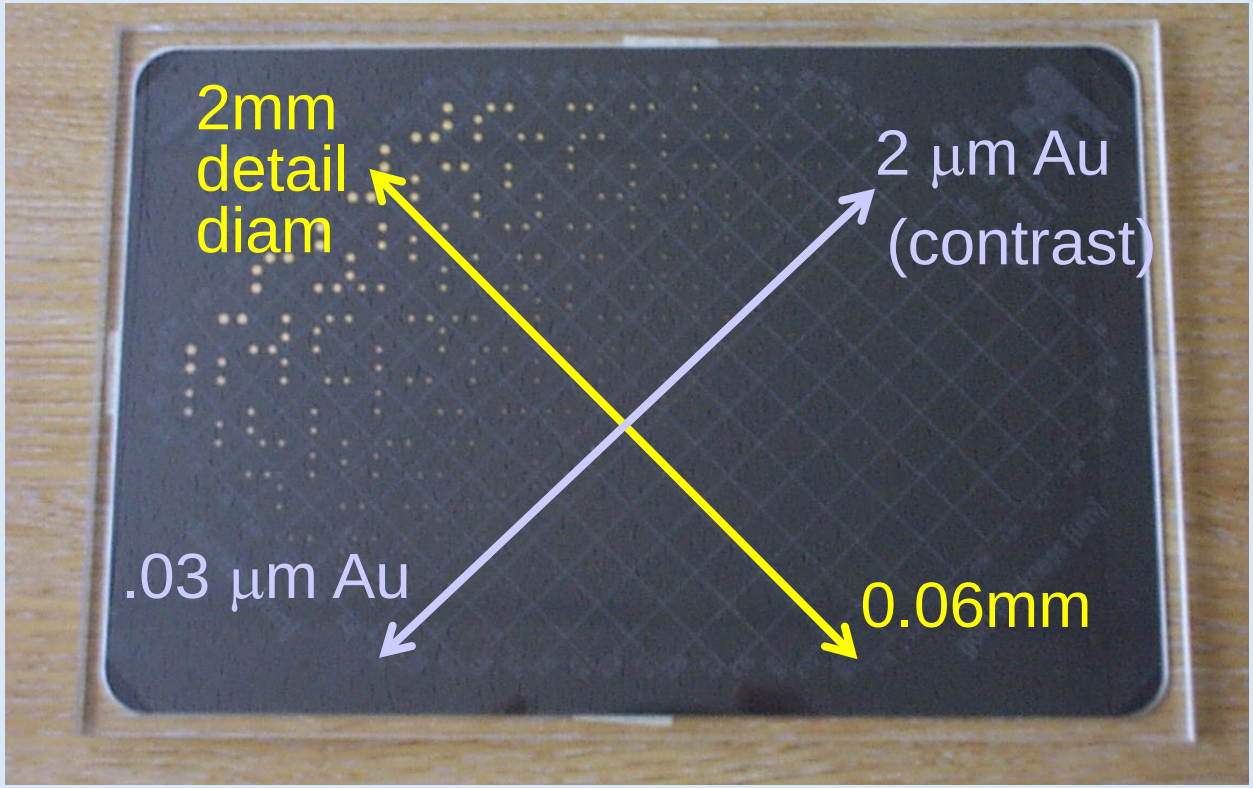
- Background
- Breast dosimetry and breast dose
- Choice of X-ray spectra
- Image quality assessment
  - Physical image quality measures
  - Quality measures using phantoms
  - Clinical measures of image quality
  - Do phantom measurements and clinical measures correlate?
- Conclusions

Contrast detail test phantom

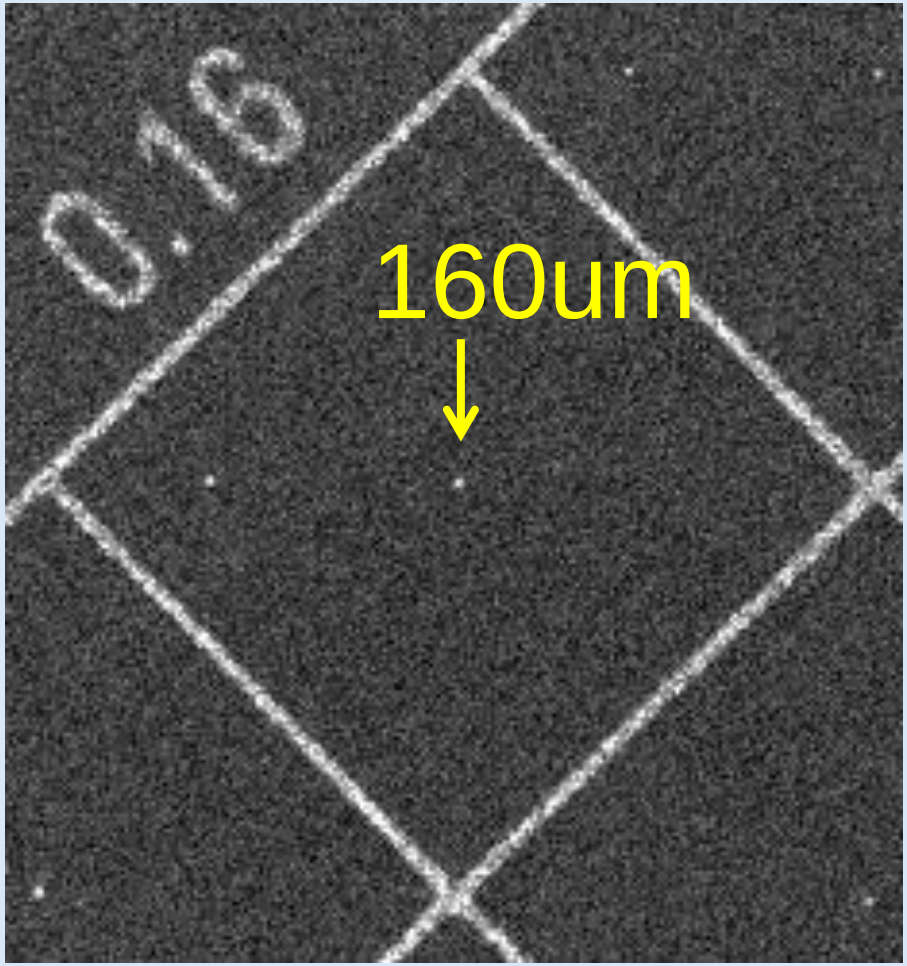
CDMAM phantom imaged with 50 mm PMMA

Can be read using CDCOM software

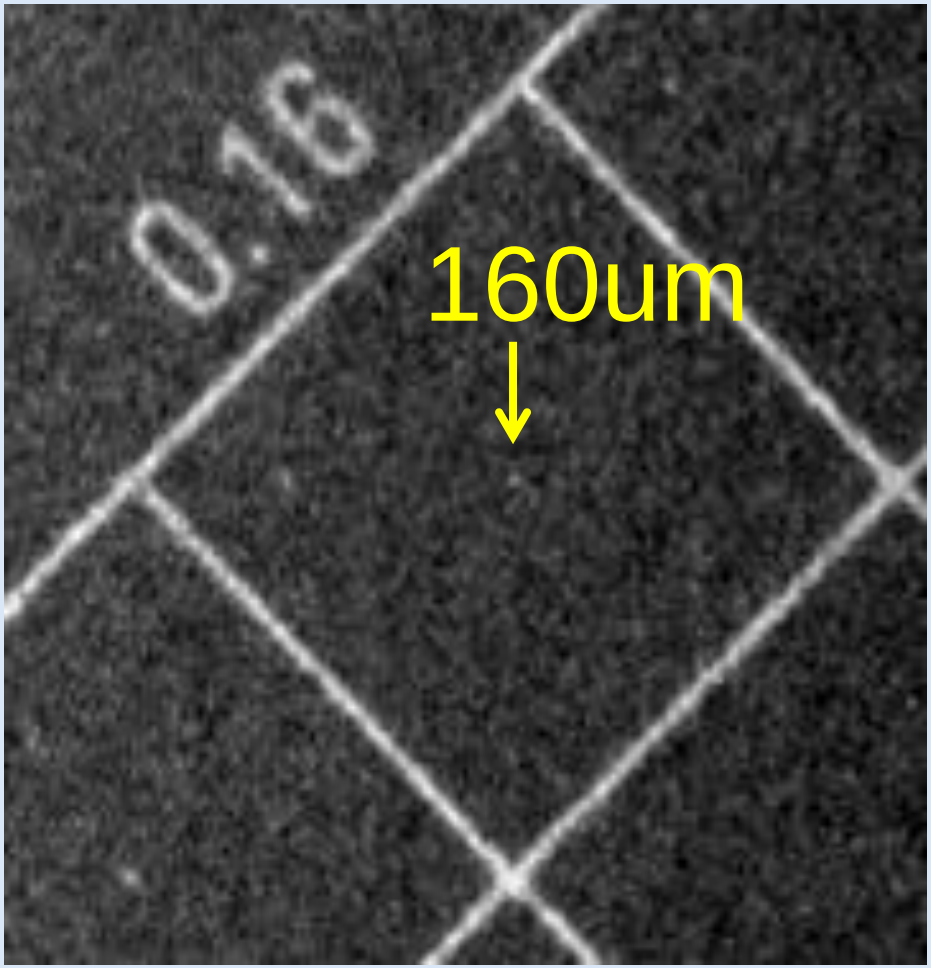
Uses unprocessed images (ideally)



Example: 160 µm detail



Good System



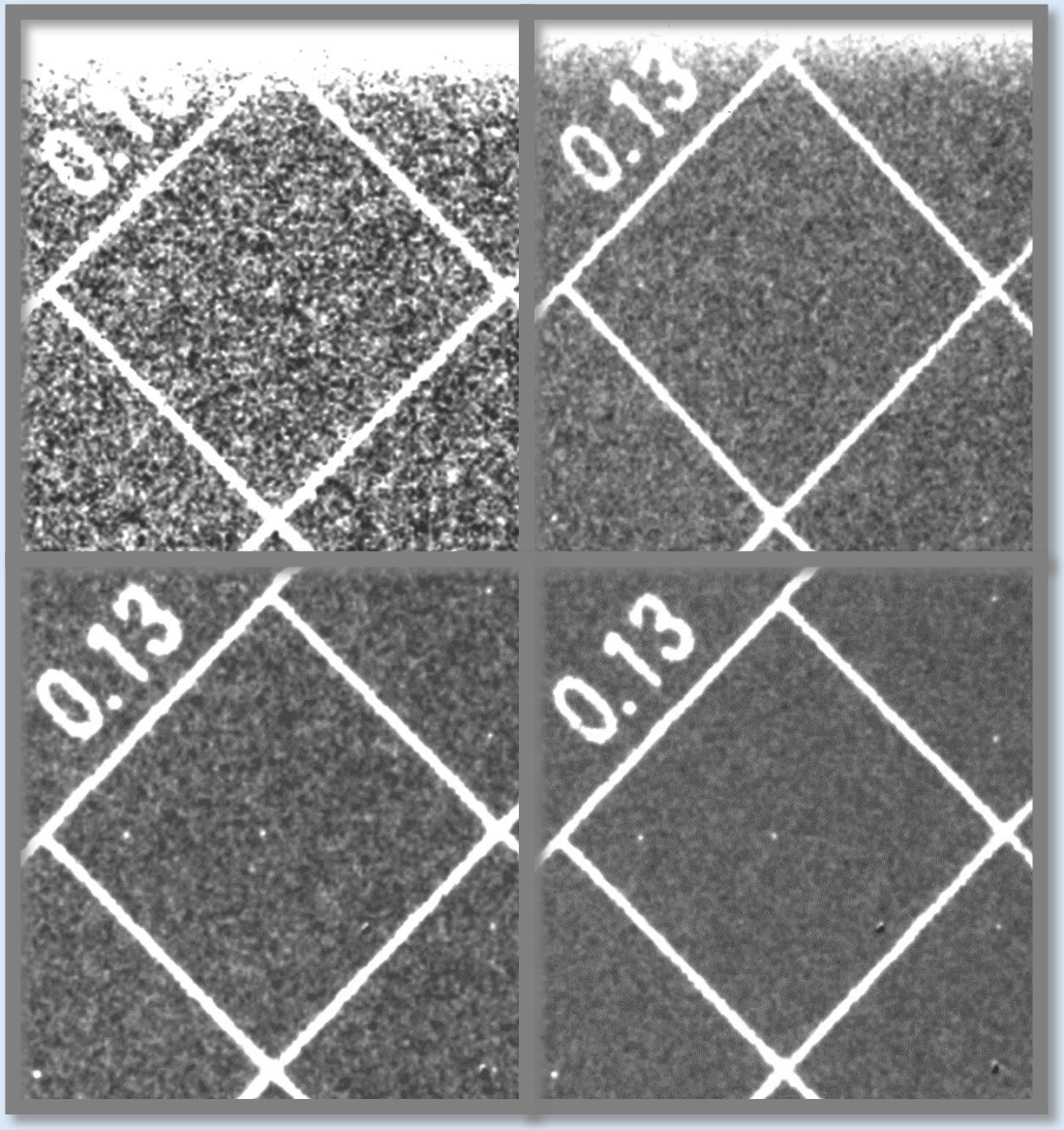
Poor System

How were detectability standards set?

- Minimum gold thickness detectable at diameters 0.1 to 2 mm
- Acceptable set so that 97.5% of systems used in UK breast screening would comply
  - **NOT VERY DEMANDING**
- Achievable values set as averages of data from established digital systems recognised to have ‘good quality’

Dose and noise

40 mAs



80 mAs

160 mAs

320 mAs

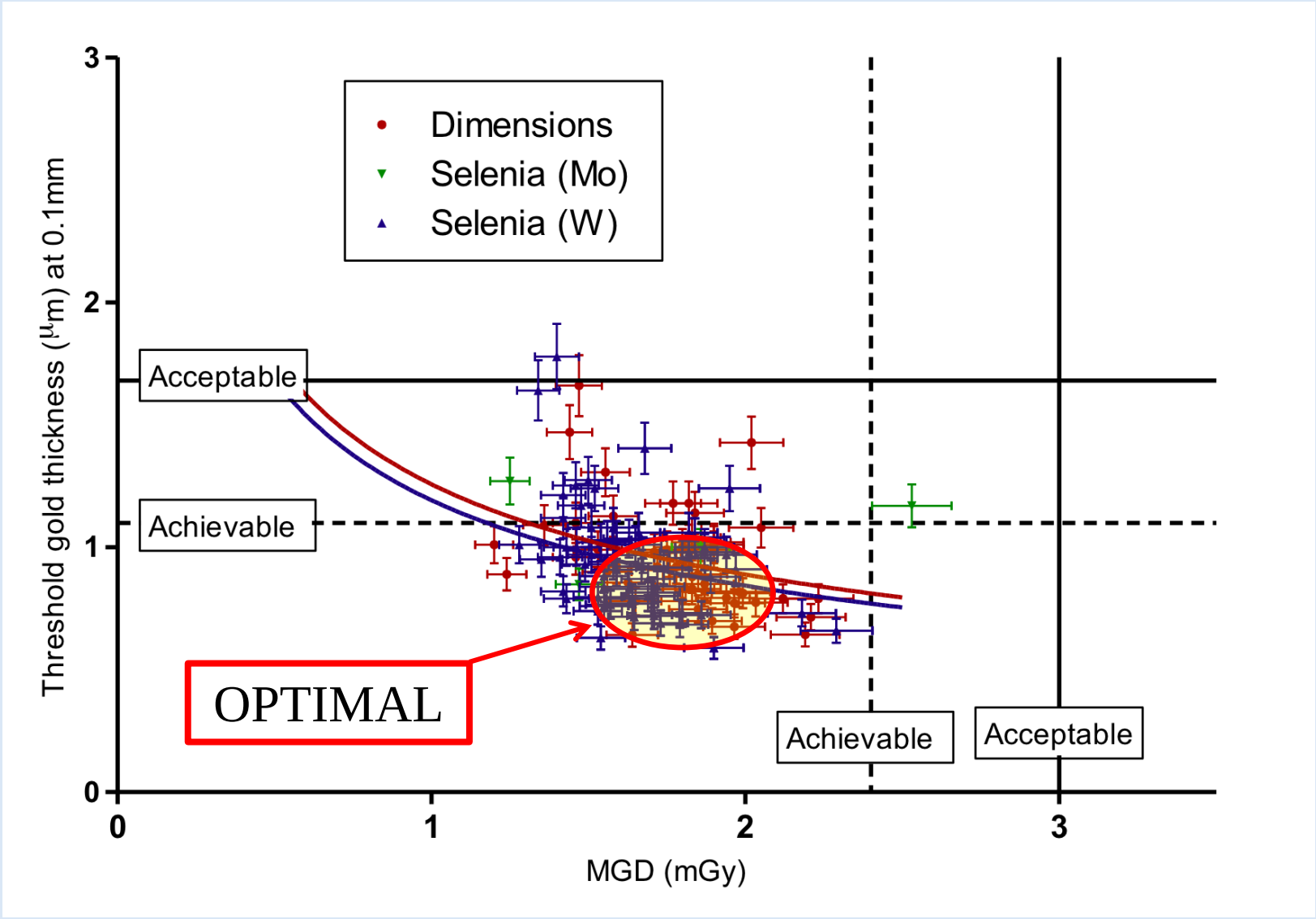
Dose required to reach minimum image quality standard (60mm thick breast)

SYSTEM	mGy (0.1 mm)	mGy (0.25 mm)
DR 1	0.60	0.67
DR 2	0.63	0.52
DR 3	0.85	0.80
DR 4	1.01	0.87
Average film-screen	1.17	1.07
CR 1	1.67	1.45
CR 2	3.46	1.49

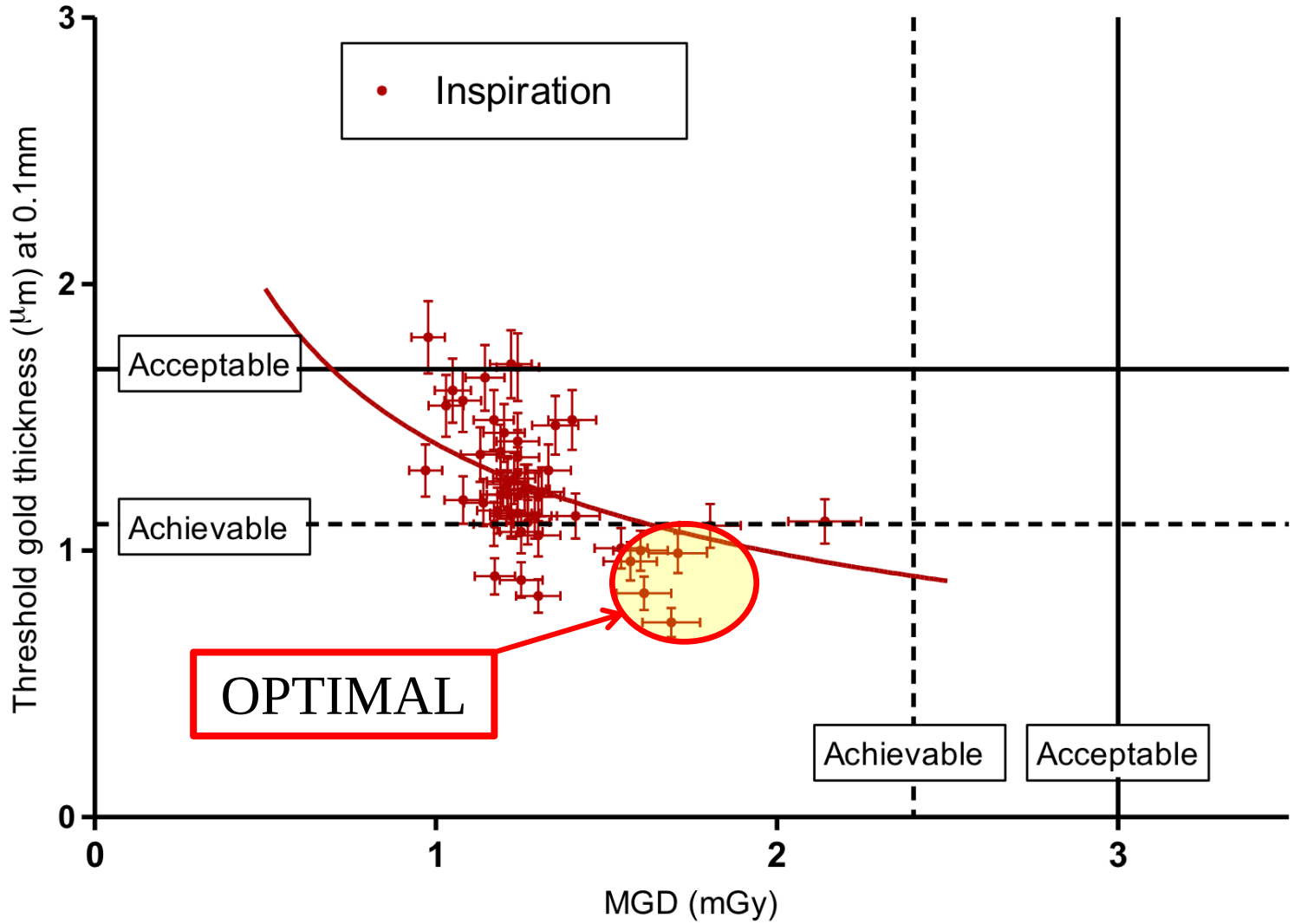
How well are systems in NHSBSP optimised?

Plot dose v image quality from data for 318 systems

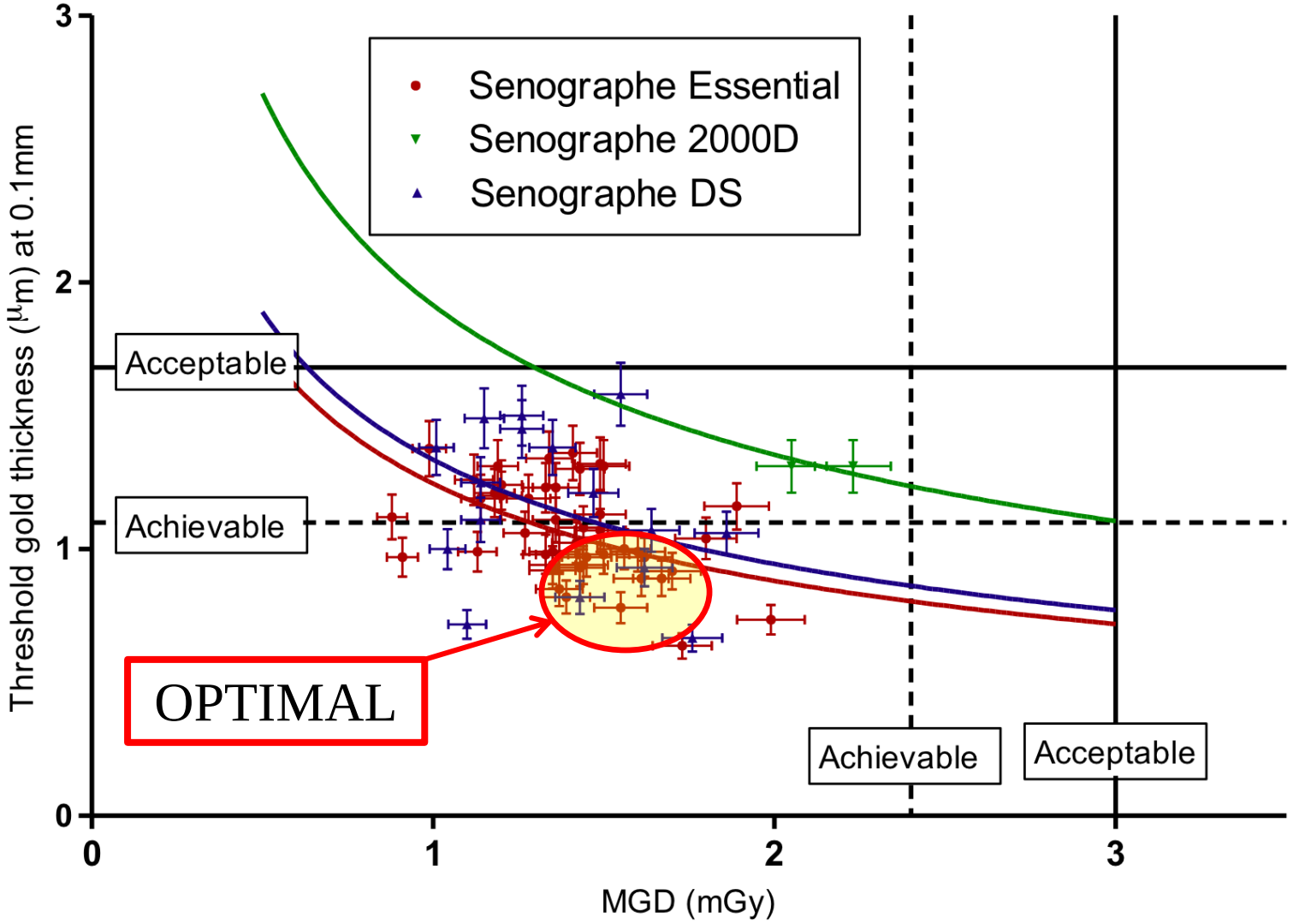
Hologic systems in NHSBSP



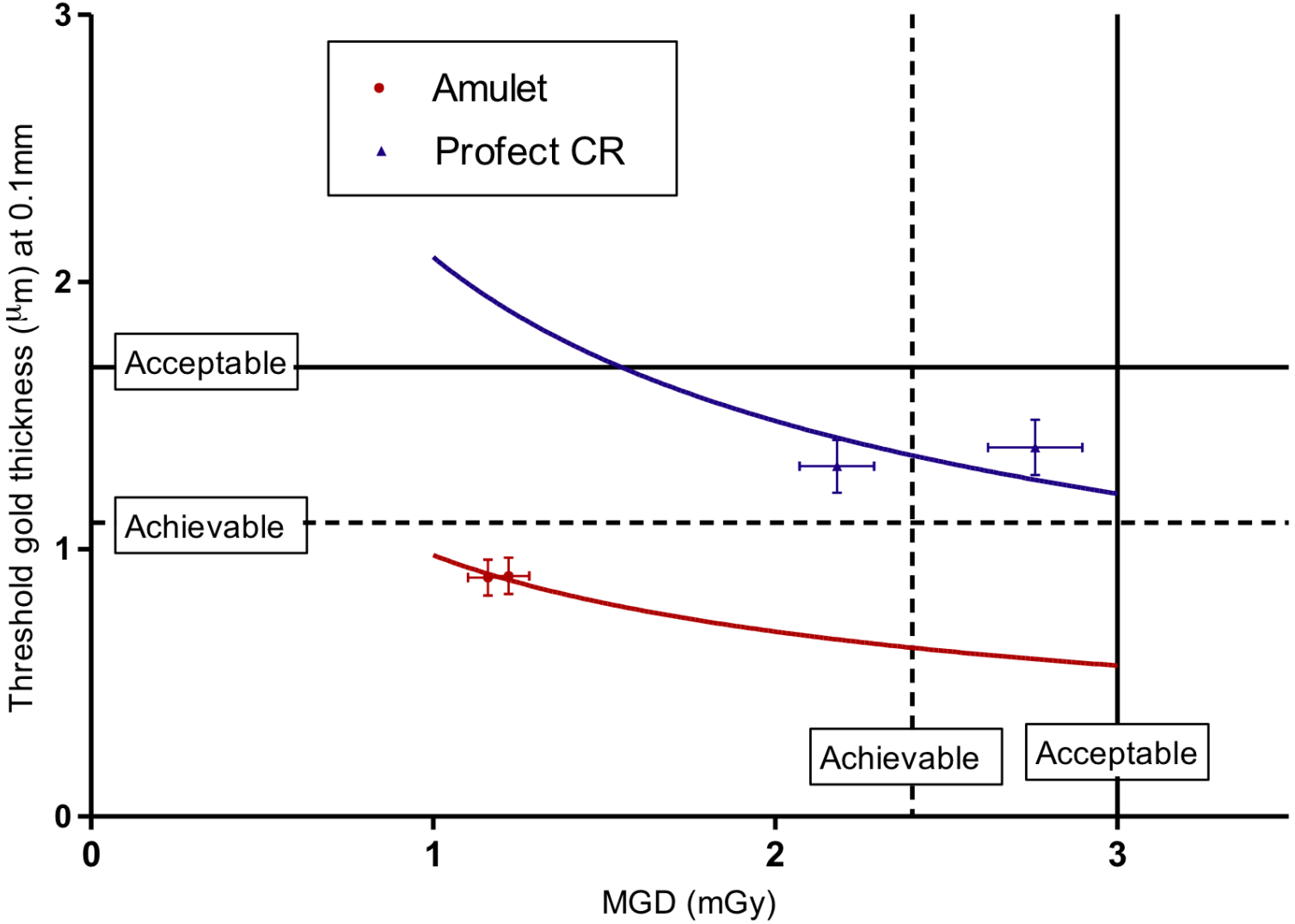
Siemens systems in NHSBSP



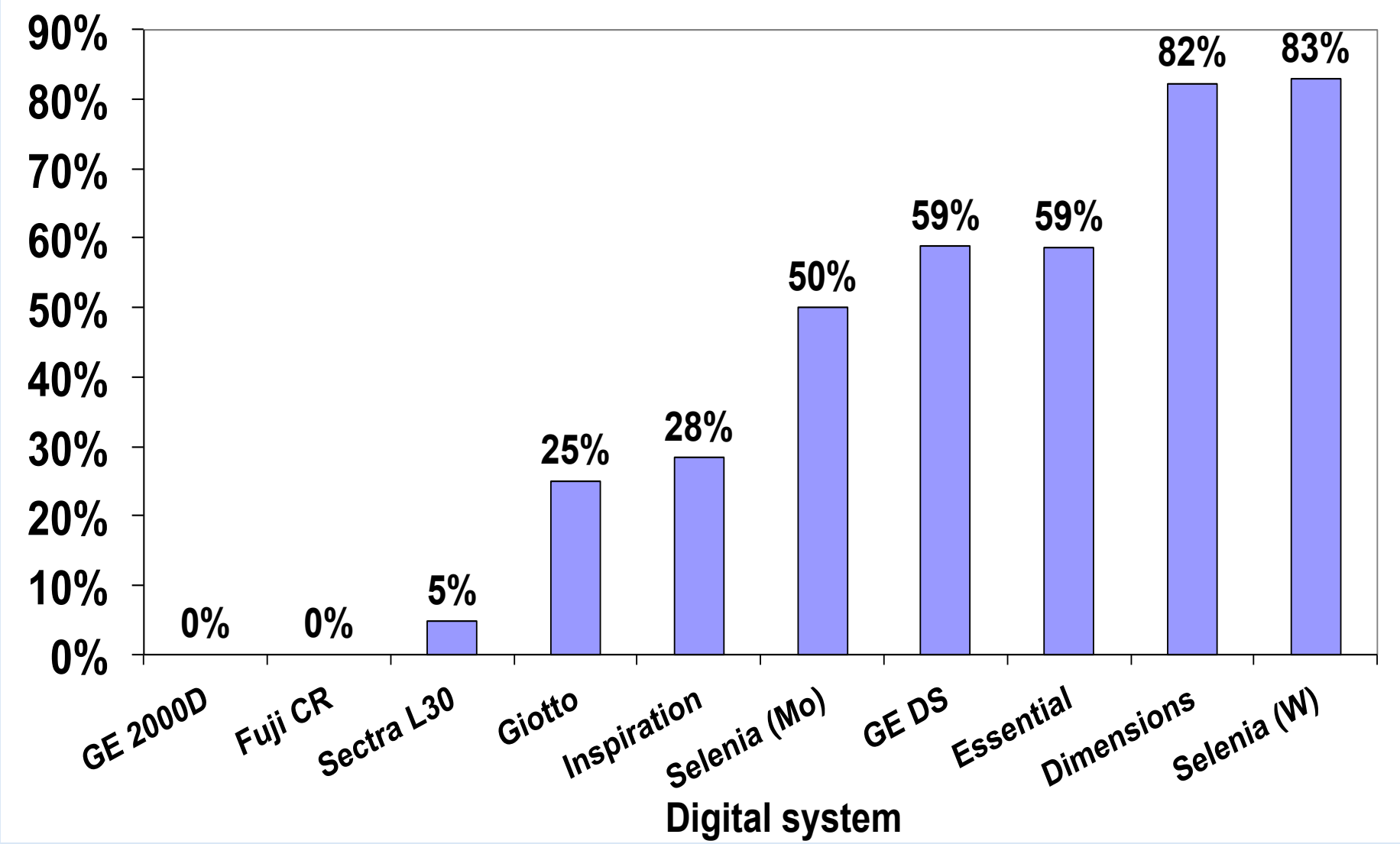
GE systems in NHSBSP



Fuji systems in NHSBSP



Proportion exceeding achievable IQ



Drawbacks of contrast detail measures

- Standards set using unprocessed images
- Phantoms have no anatomical background
- Details may not be clinically realistic

Outline

- Background
- Breast dosimetry and breast dose
- Choice of X-ray spectra
- Image quality assessment
  - Physical image quality measures
  - Quality measures using phantoms
  - Clinical measures of image quality
  - Do phantom measurements and clinical measures correlate?
- Conclusions

Clinical measures of image quality 1

Results from breast screening with different equipment

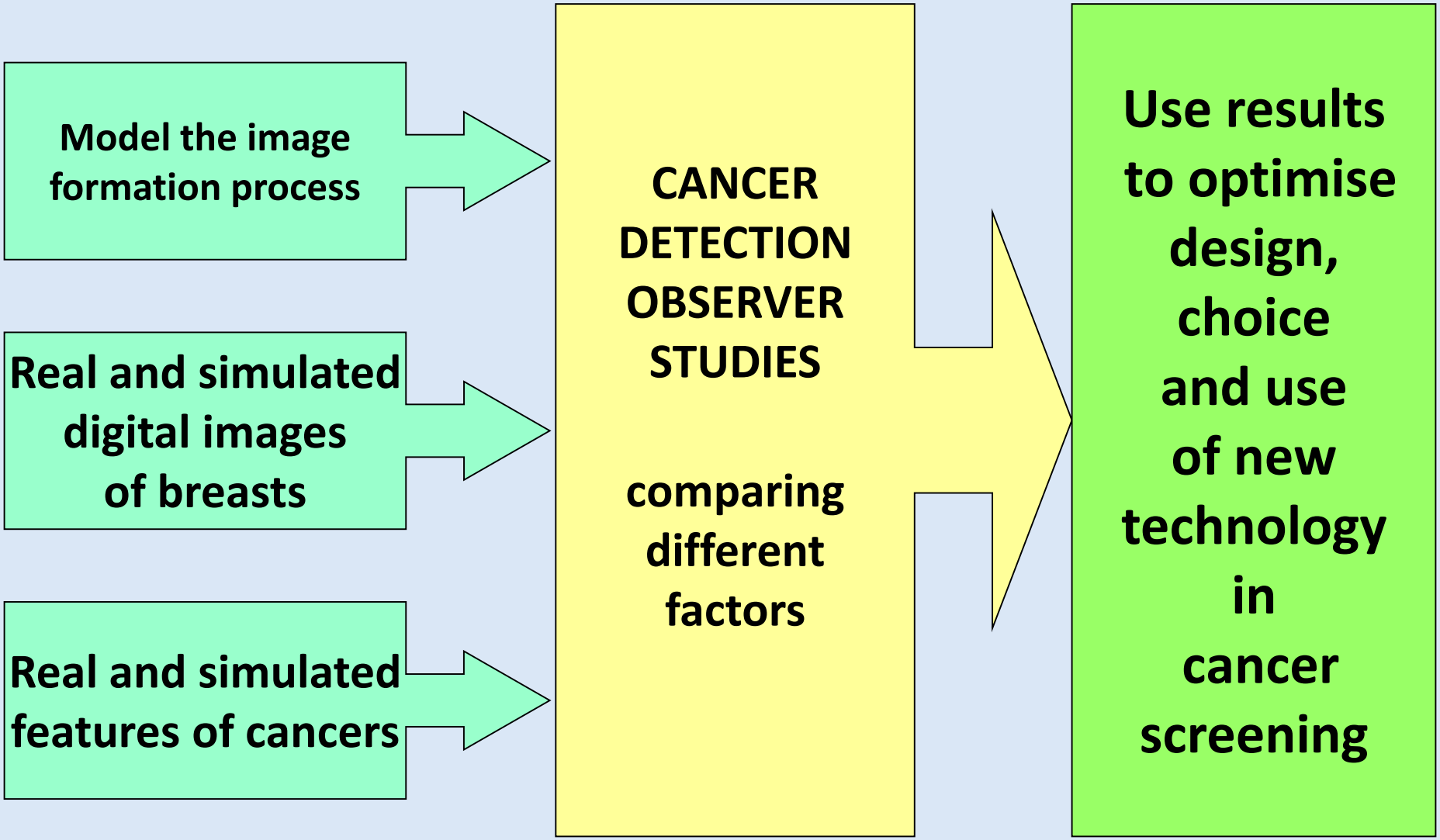
Differences in cancer detection using CR compared with DR in screening programs

Study	DCIS	Invasive cancer	Combined lesions	CR/DR dose
Belgium 2013	**-27%**	**+2%**	**-3.5%**	**1.6**
Canada 2013	**-50%**	**-18%**	**-29%**	**1.17**
France 2014	**-53%**	**-15%**	**-23%**	**NA**

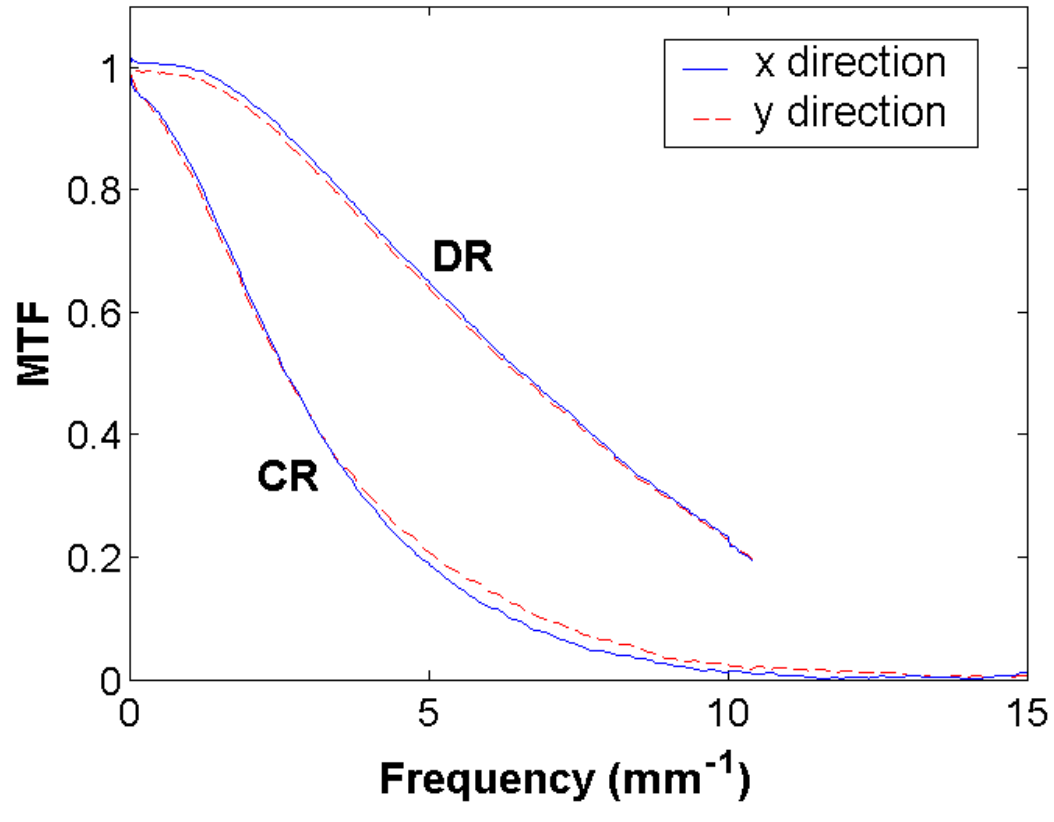
Clinical measures of image quality 2

- Difficult to make direct comparison of systems for breast screening by imaging the same breast with each system
  - additional dose
  - low incidence of cancer
  - breast positioning may differ for each system
- Use virtual clinical trials
  - real (or simulated) images
  - each image modified mathematically to simulate different systems
  - can use real lesions or inserted simulated lesions
- **Results of 3 virtual clinical trials from OPTIMAM**

Our Experimental Approach



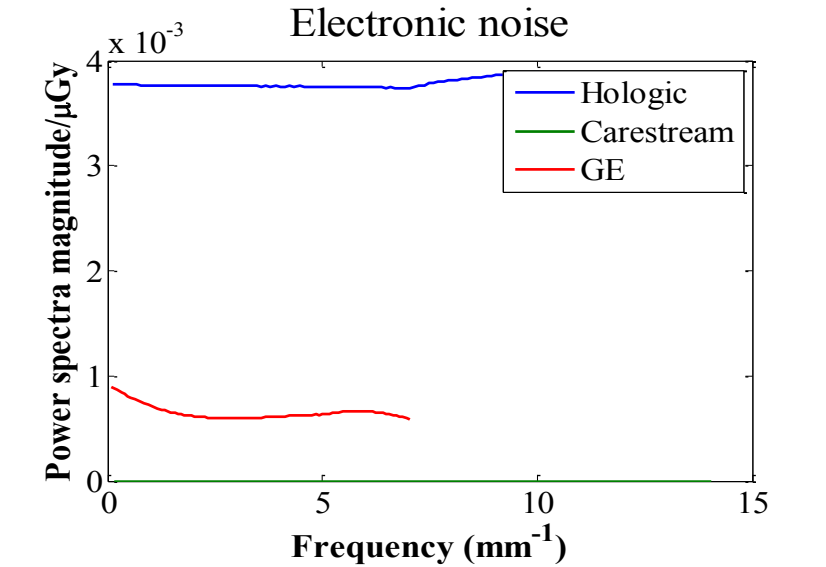
Modifying image appearance 1

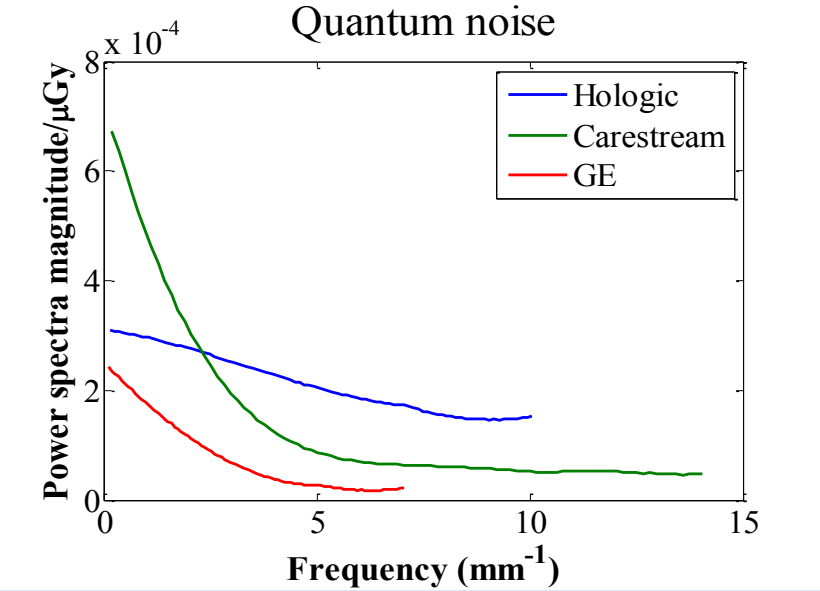


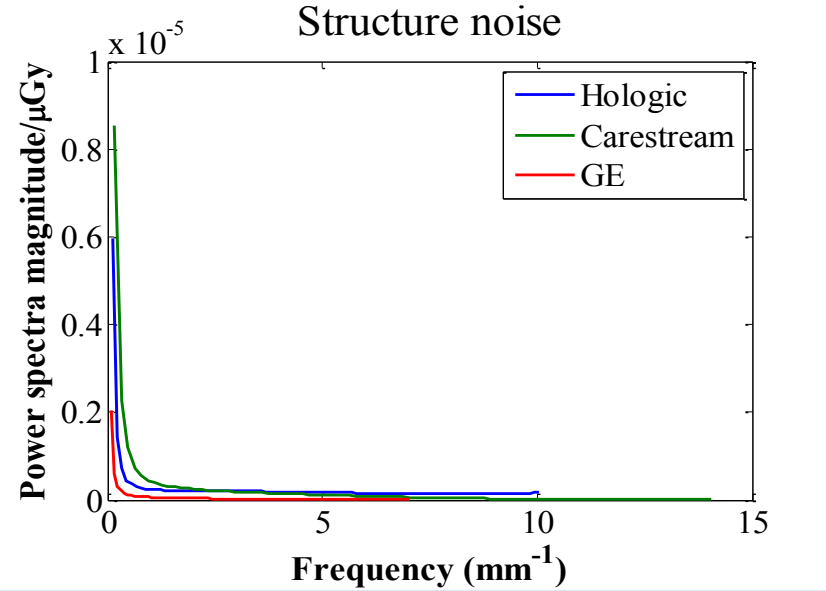
Adjust image unsharpness using MTF measurements

Modifying image appearance 2

Measure and understand noise power behaviour with dose and spatial frequency





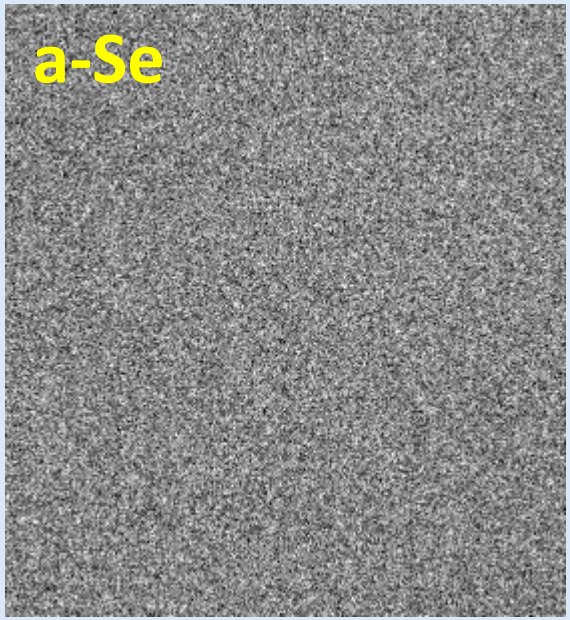


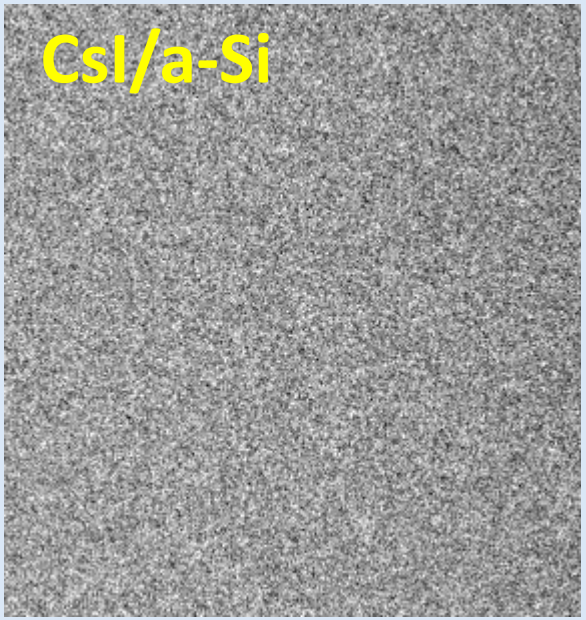
A Mackenzie et al. Medical Physics 2012

NCCPM

Adapt images by adding noise

- Calculate extra noise
  - For dose change
  - For differences between detectors
- Create noise images to add to real image
  - On pixel-by-pixel basis to account for different signal levels





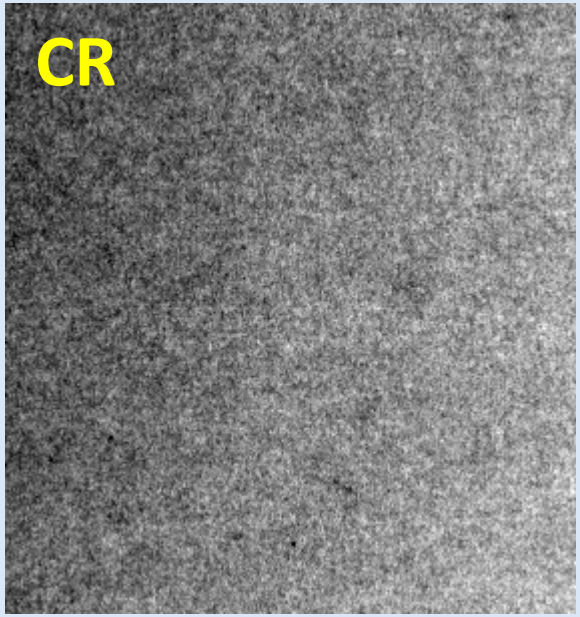
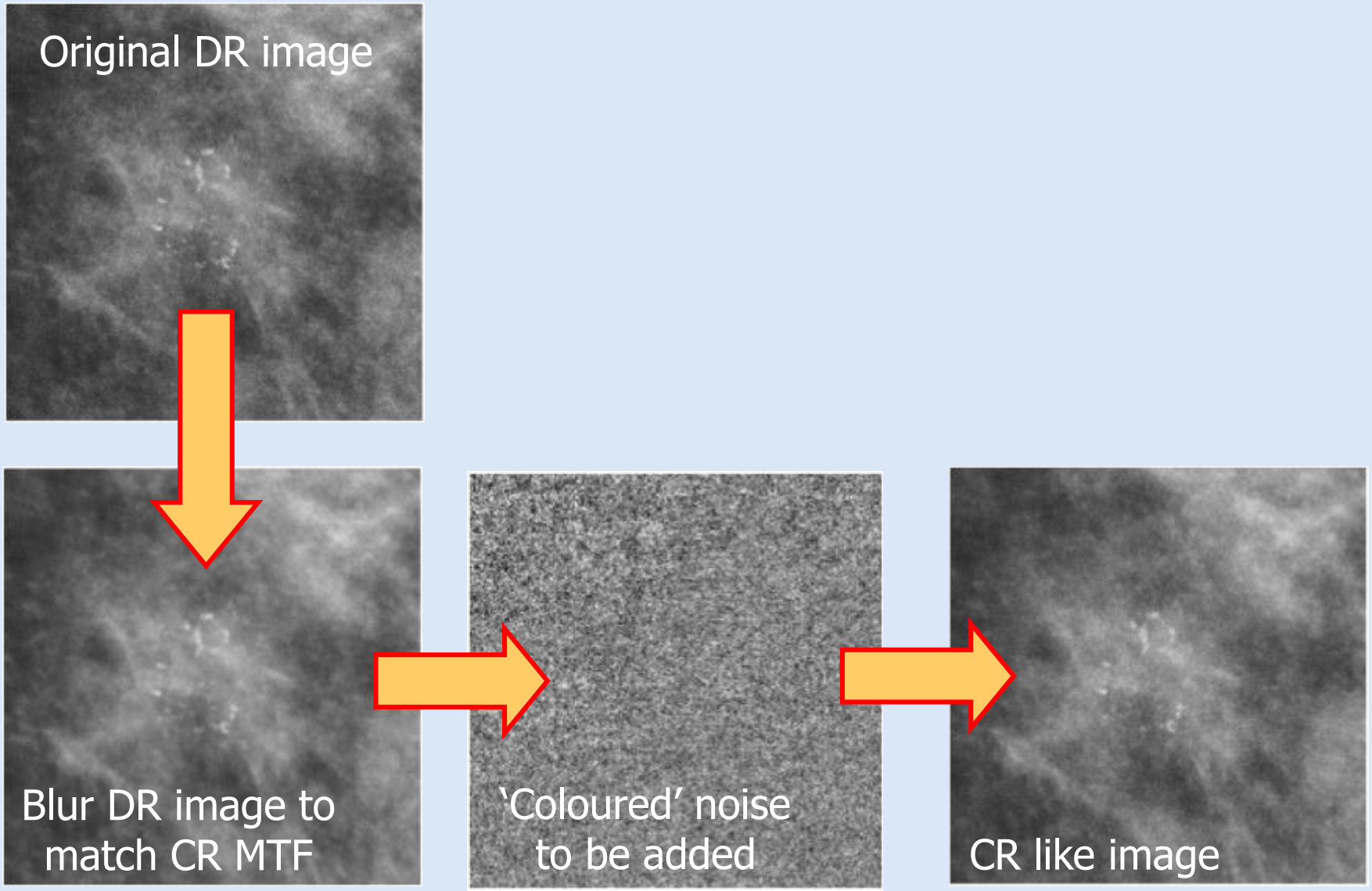


Image quality modification (DR to CR)



OBSERVER STUDY 1

Calcification detection at different image qualities

Observer study on calcification detection

81 normal cases

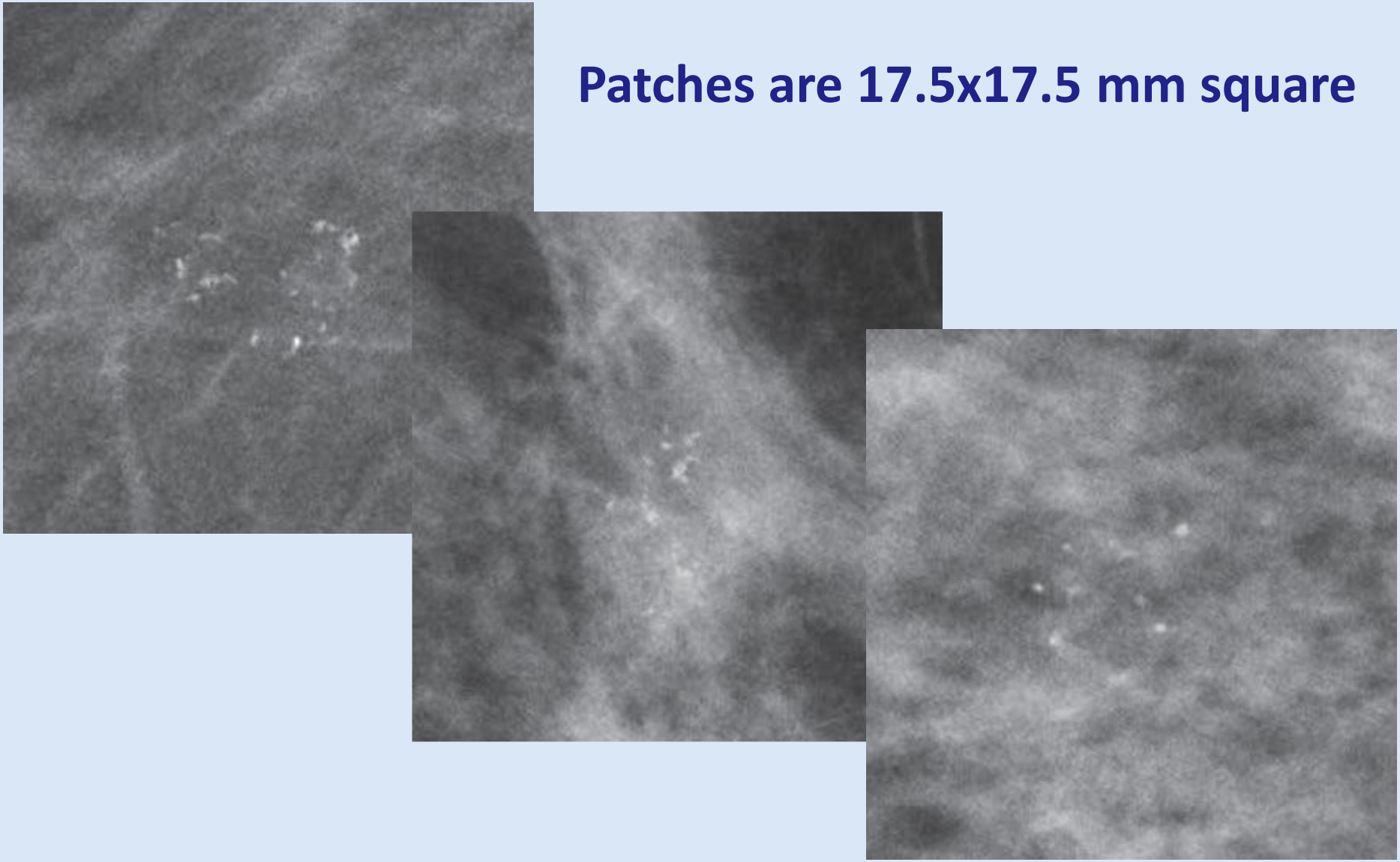
81 abnormal cases with 113 simulated subtle clusters

6 image quality levels

7 observers

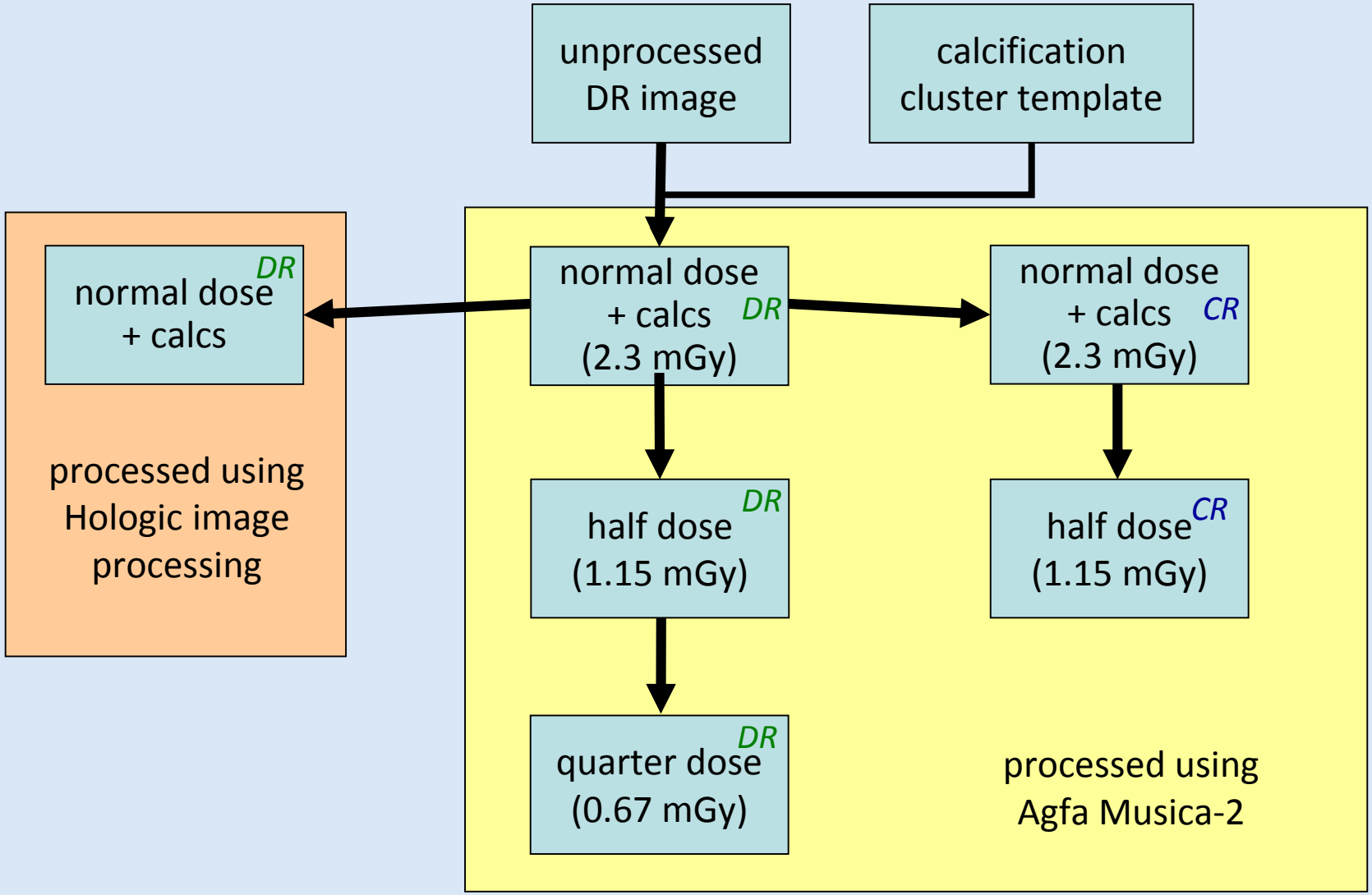
6804 image readings

Simulated calcifications



Patches are 17.5x17.5 mm square

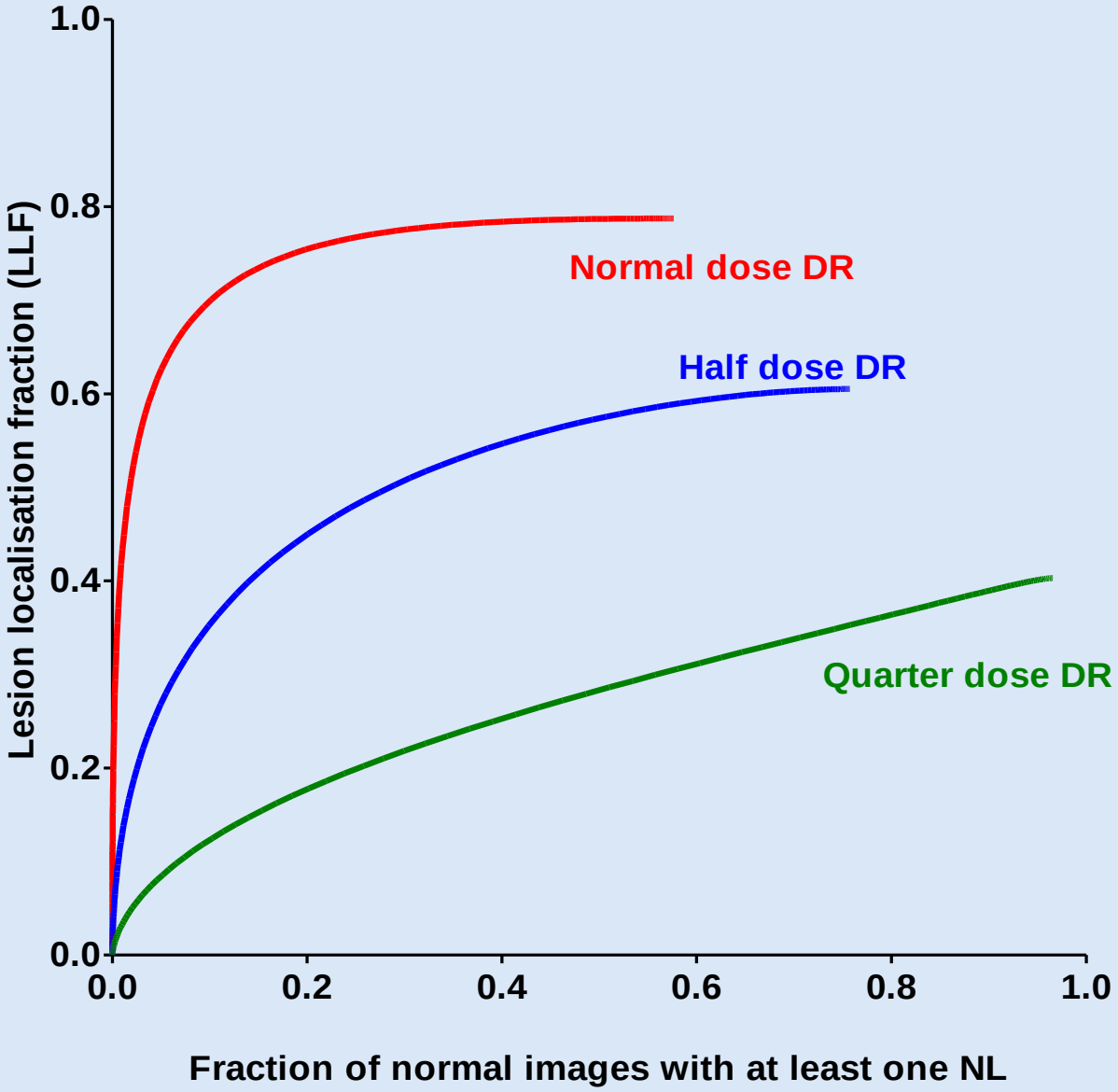
Each DR image used to create additional images at 6 different IQ levels



L Warren et al, Medical Physics 2012

NCCPM

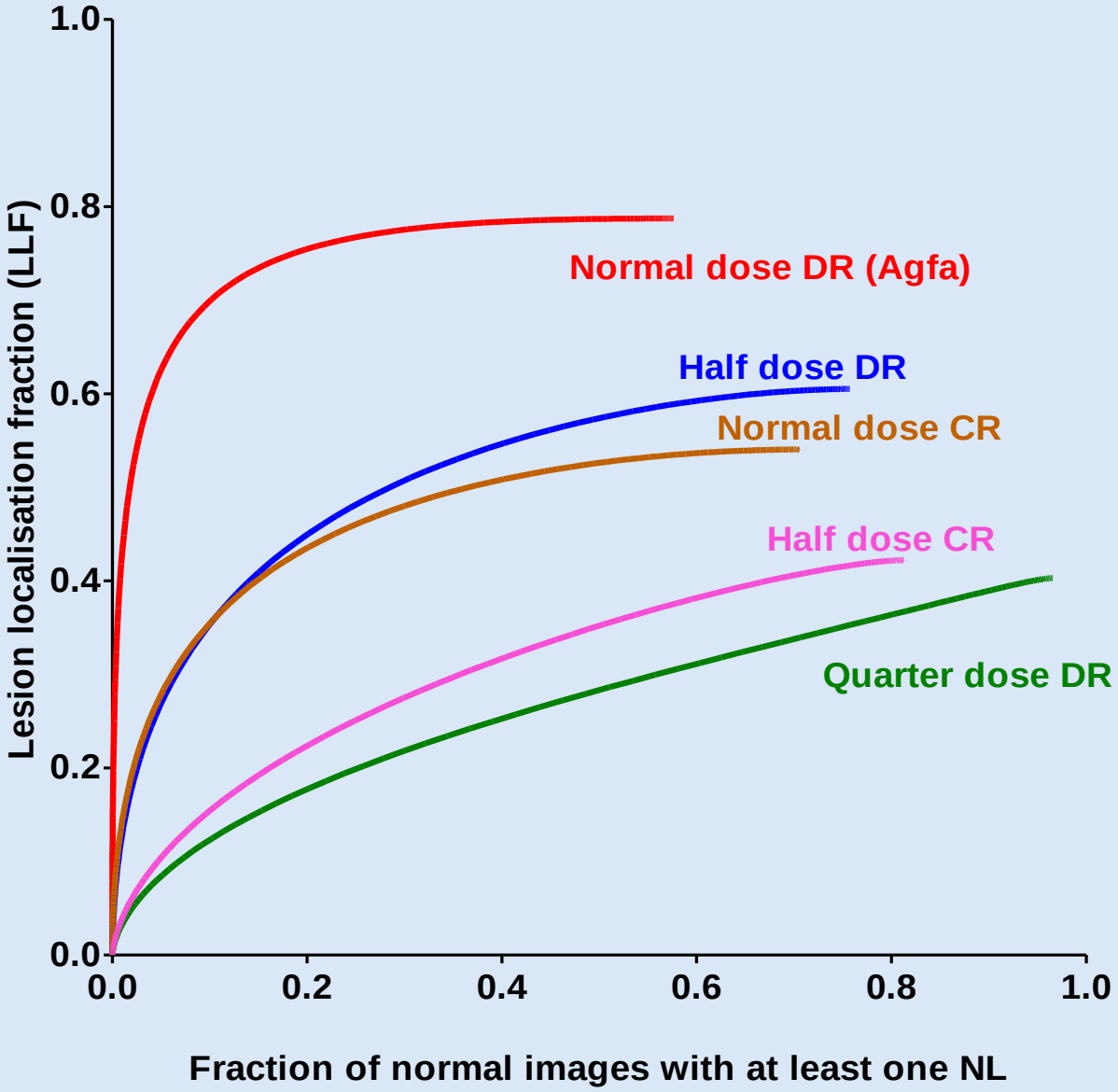
AFROC Curves for IQ levels
(fitted to average of 7 observers)



L Warren et al, Medical Physics 2012

NCCPM

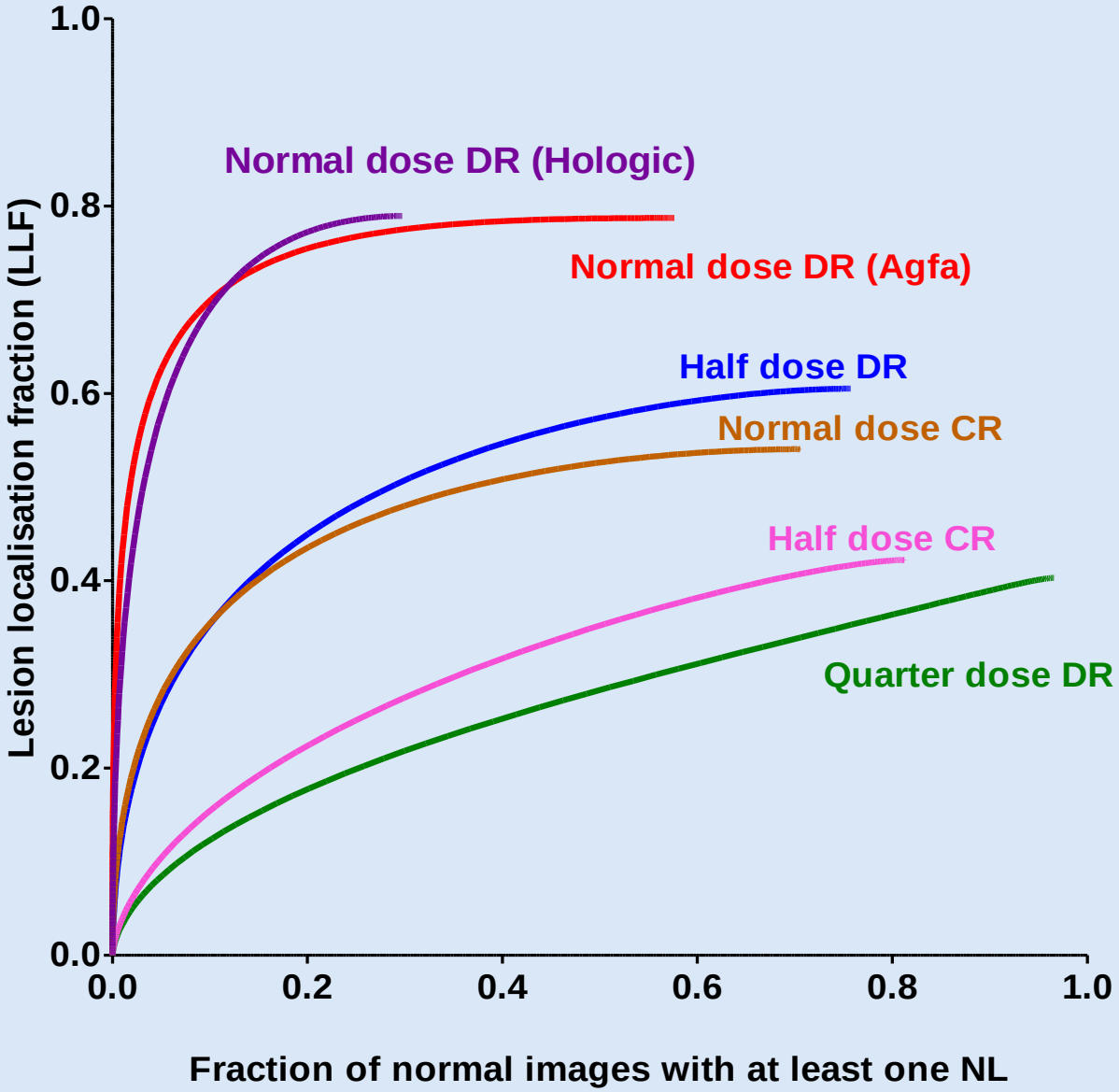
AFROC Curves for IQ levels
(fitted to average of 7 observers)



L Warren et al, Medical Physics 2012

NCCPM

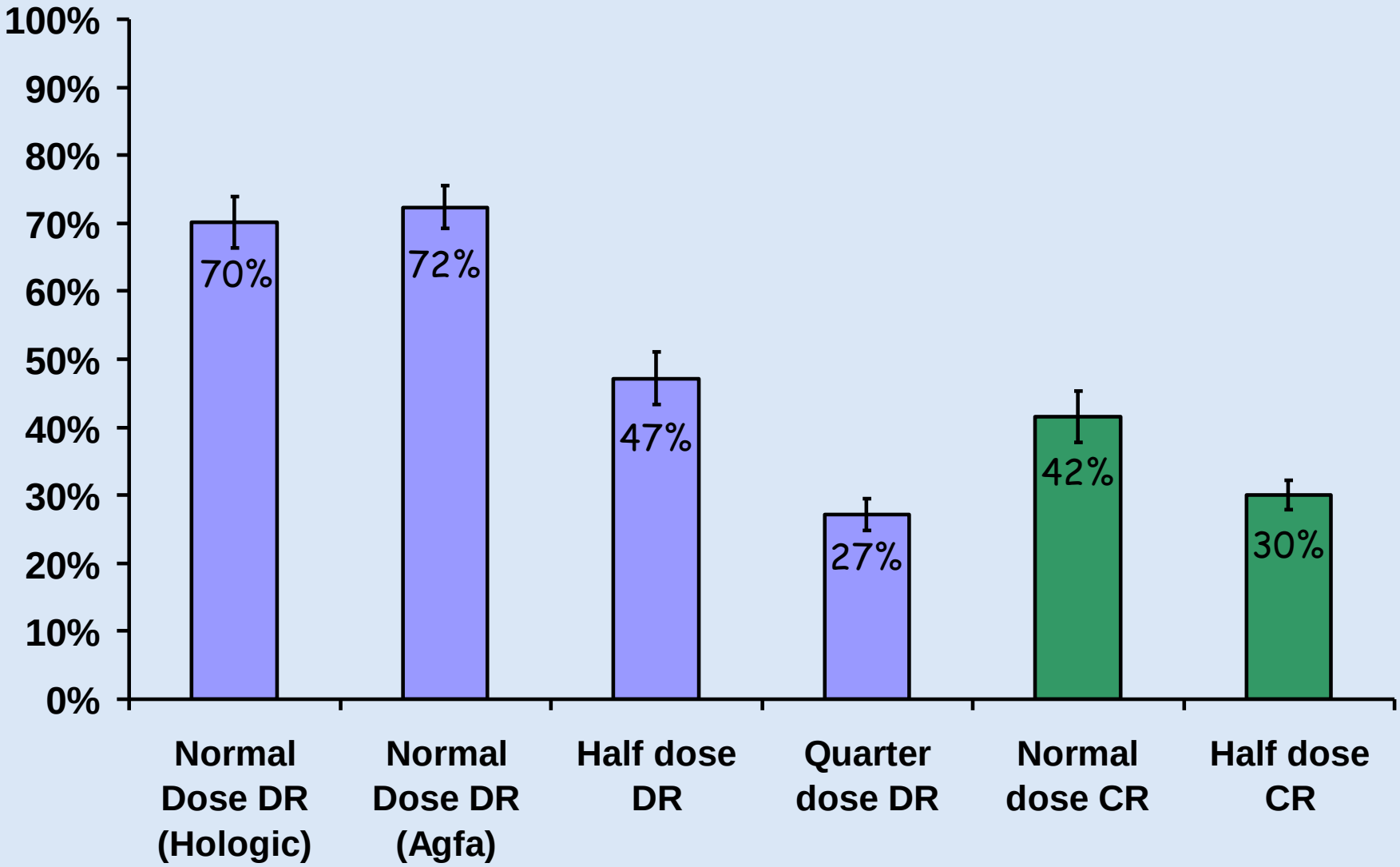
AFROC Curves for IQ levels
(fitted to average of 7 observers)



L Warren et al, Medical Physics 2012

NCCPM

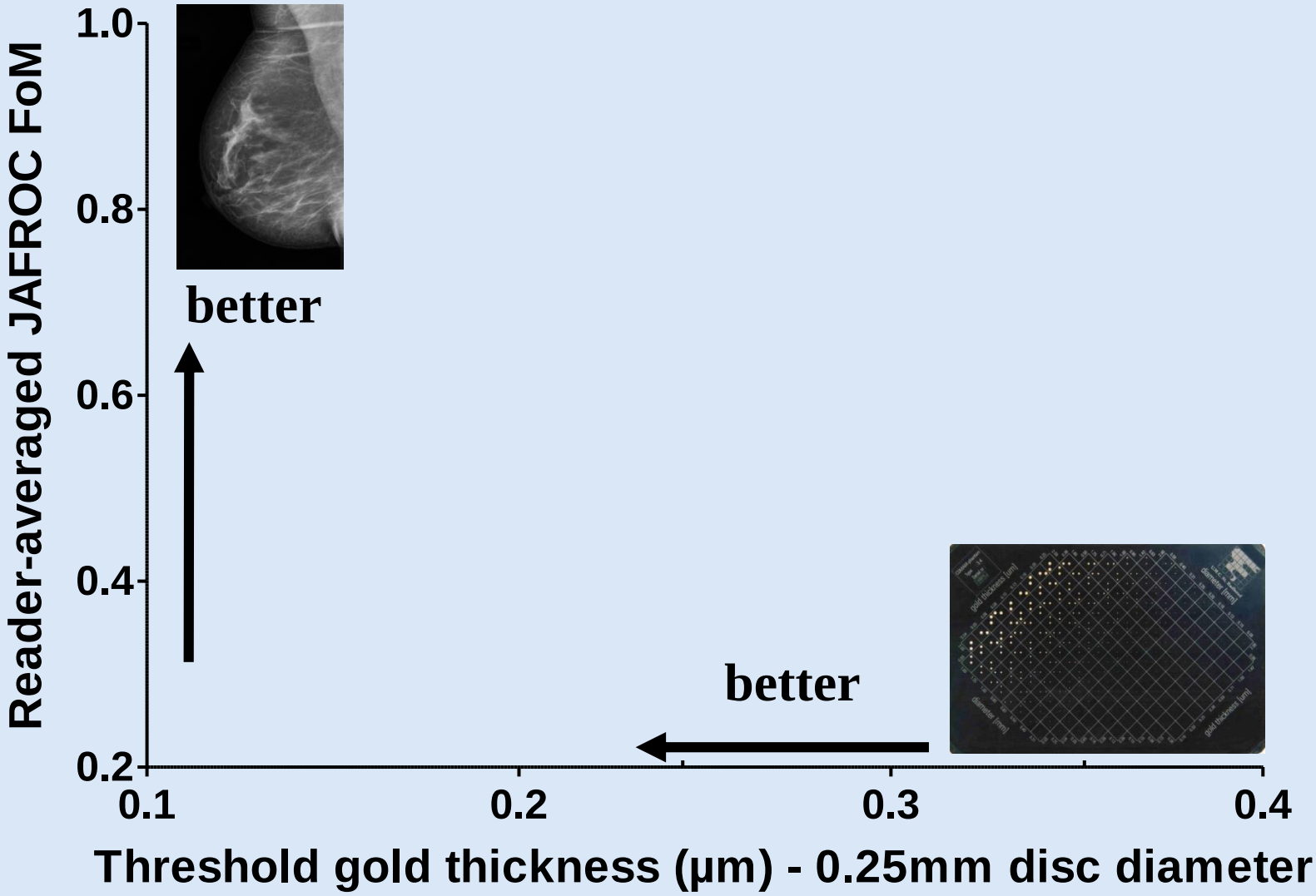
Lesion sensitivity at 0.1 FP per image



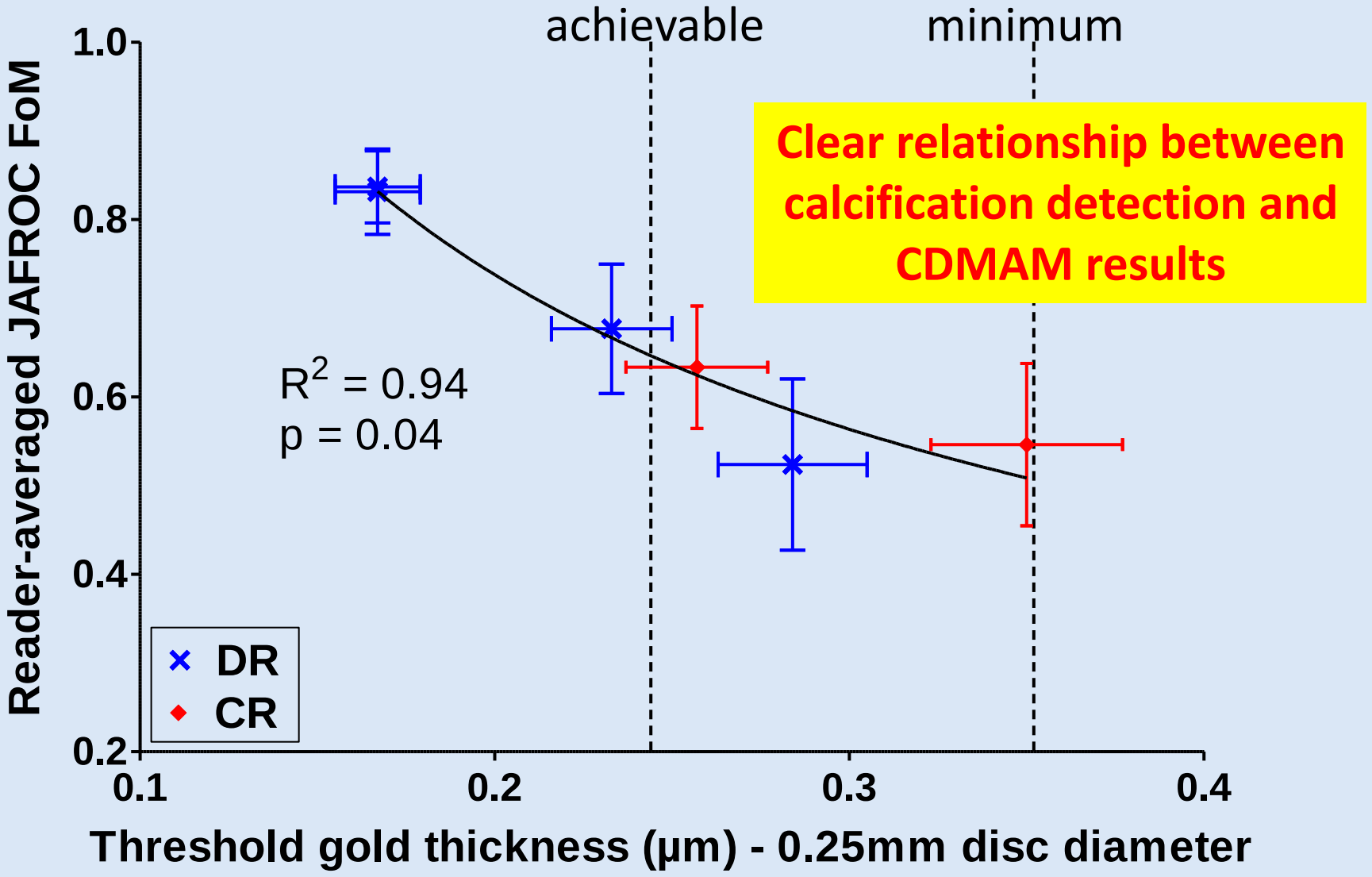
L Warren et al, Medical Physics 2012

NCCPM

Does CDMAM test object predict calcification detection?



Does CDMAM test object predict calcification detection?



Conclusions from Observer Study 1

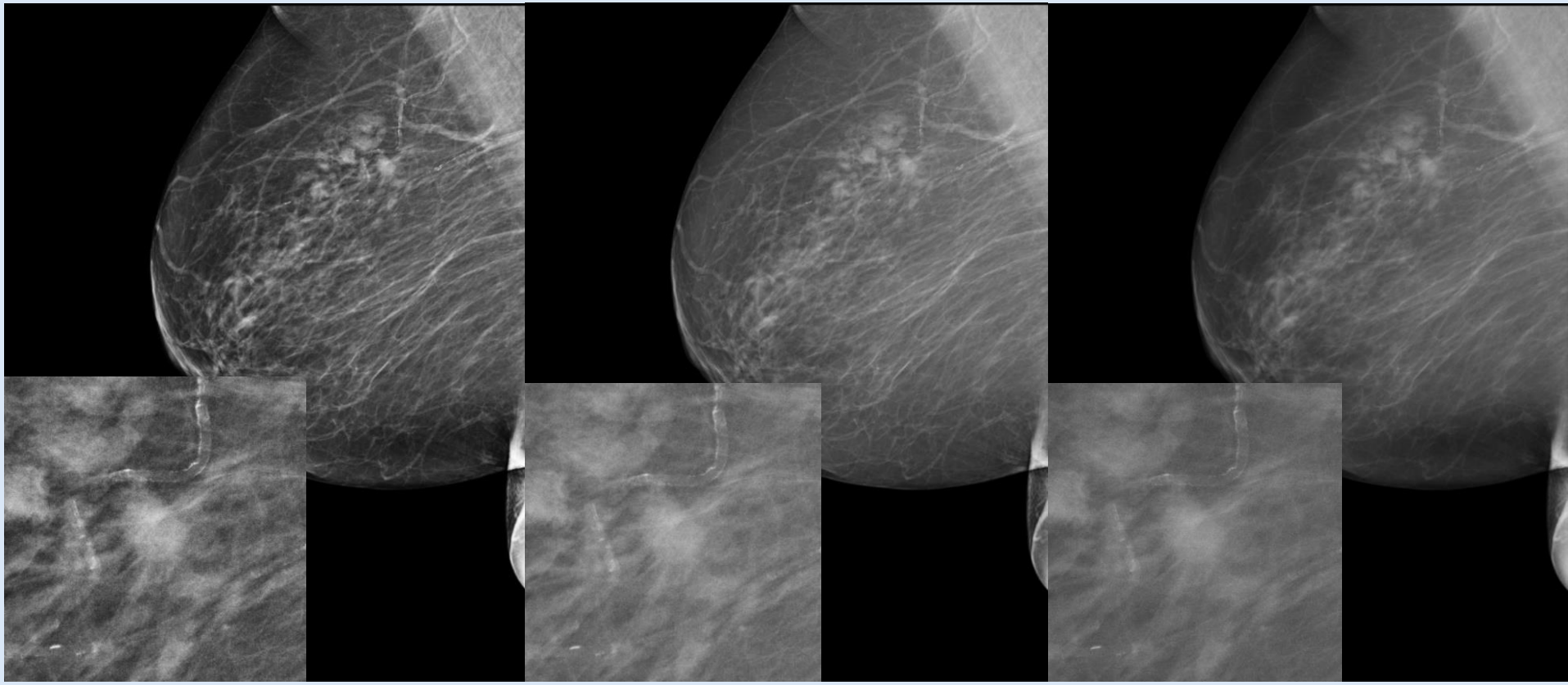
1. Threshold gold thickness using the CDMAM phantom is a good predictor of detection of clinical calcification
2. EU guidelines for IQ are clinically relevant
3. Traditional CR systems poorer than DR for calcification detection even at the relatively high dose level used (2.1 mGy, 50-60 mm breast).
4. Good IQ is important for calcification detection and systems should exceed achievable levels in EU guidelines. Acceptable level is too low
5. Image processing investigated had little effect on calcification detection

OBSERVER STUDY 2

Effect of image processing on detection of calcification and non-calcification lesions

Observer study on image processing

Hologic software



Standard version **Low contrast version** **Simulated screen-film**

L Warren et al, AJR 2014

NCCPM

Observer study on image processing

80 normal cases

80 cases with simulated subtle clusters

- avoids selection bias

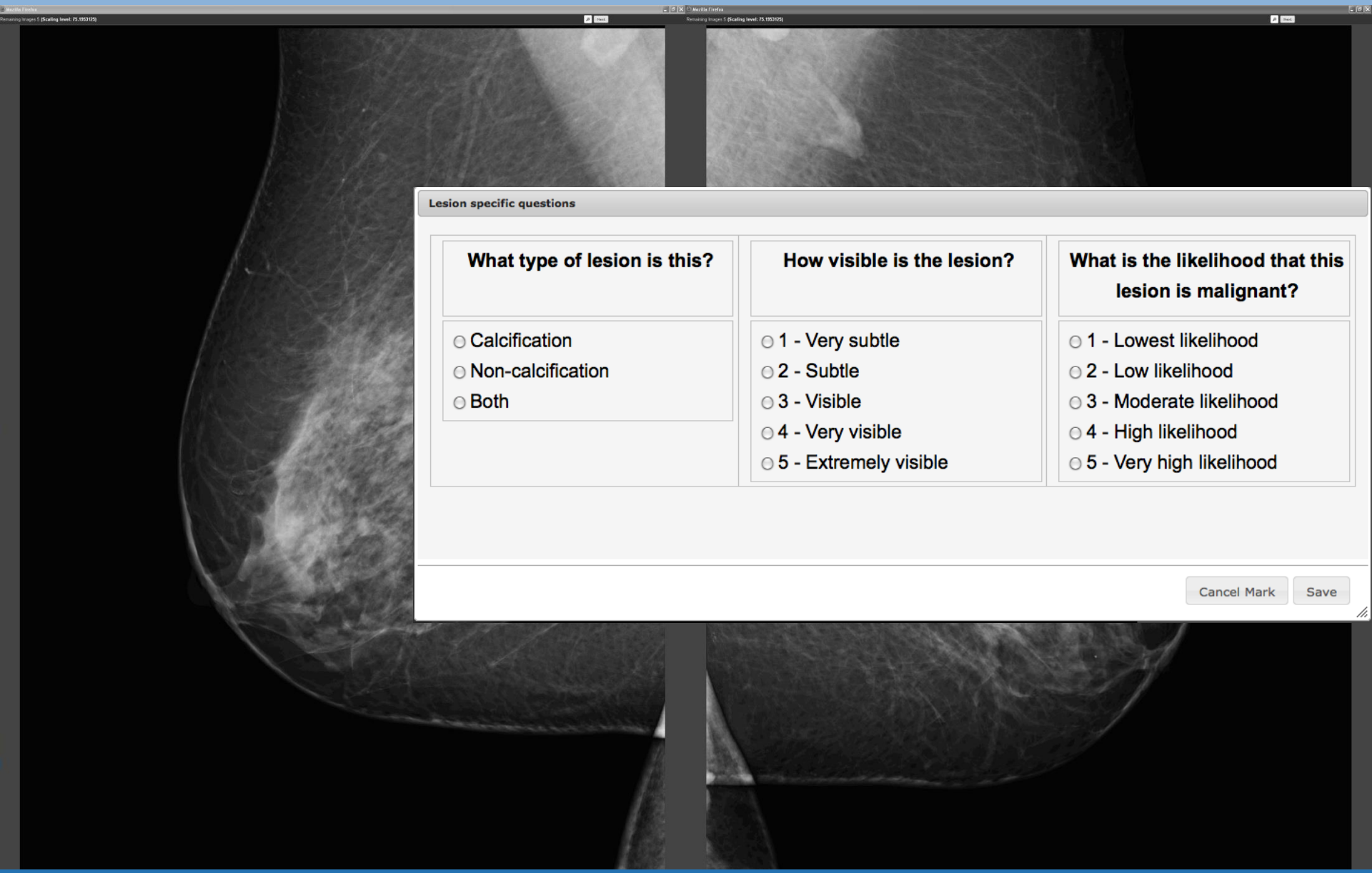
80 cases with malignant non-calcification lesions (subtle or very subtle)

30 cases of biopsy proven benign lesions

3 types of image processing

7 observers

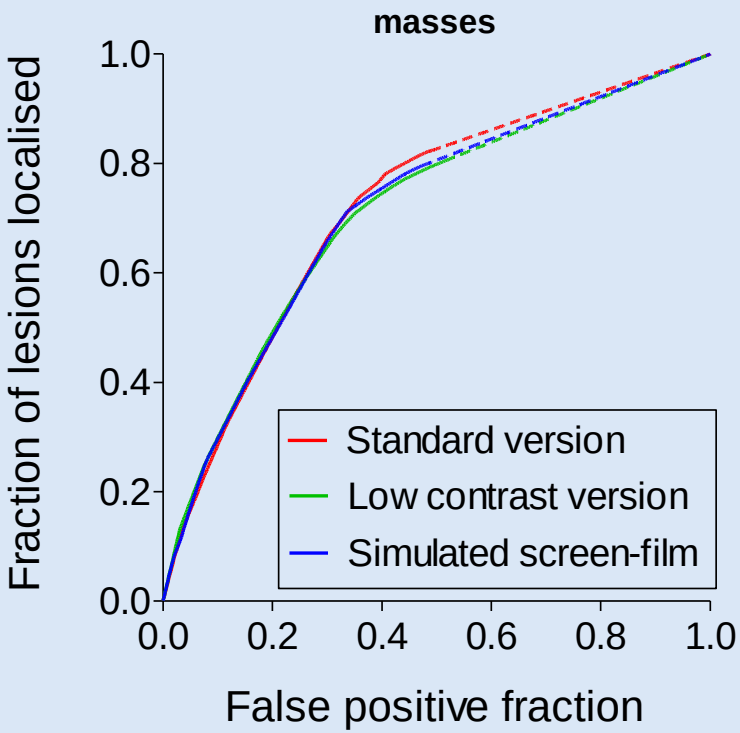
Image processing comparison



L Warren et al, AJR 2014

NCCPM

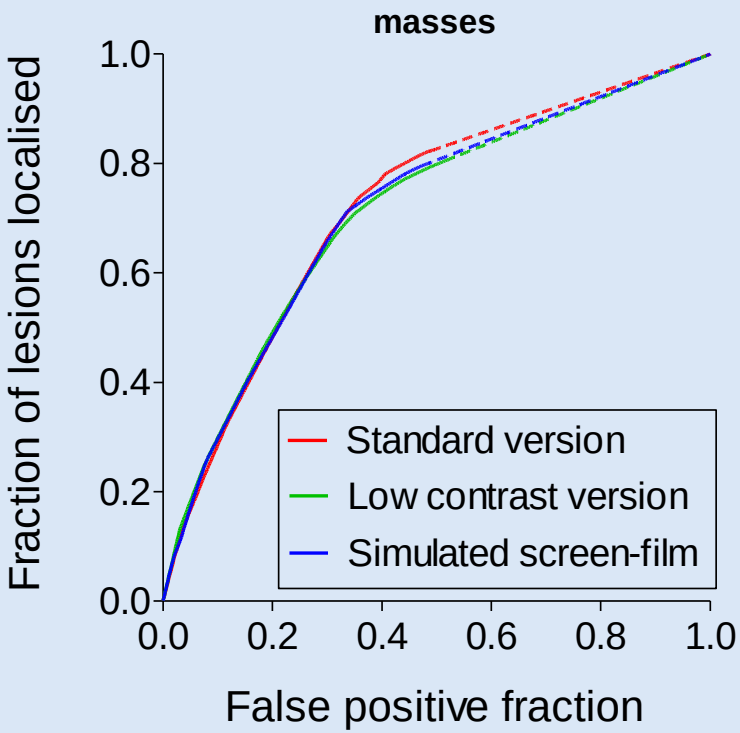
Image processing comparison

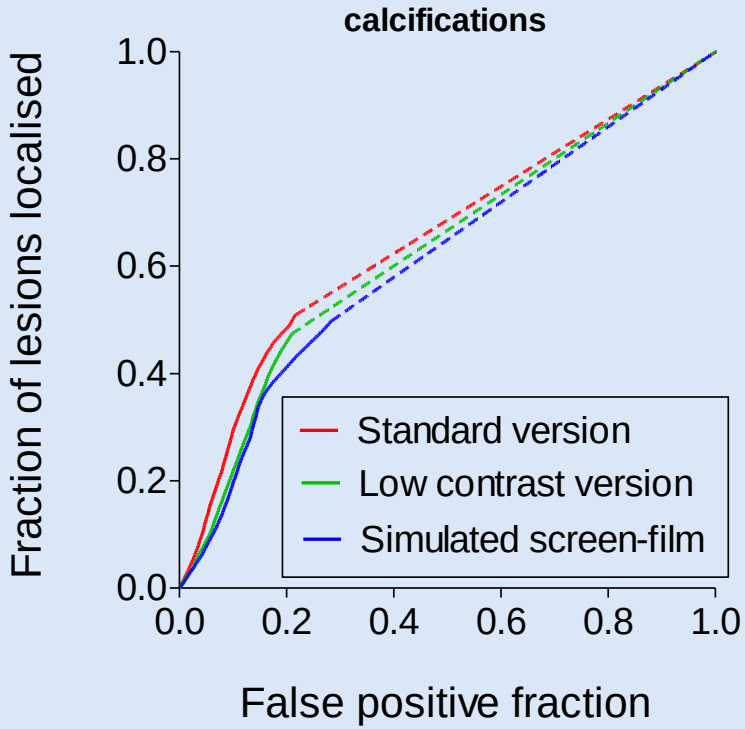


L Warren et al, AJR 2014

NCCPM

Image processing comparison

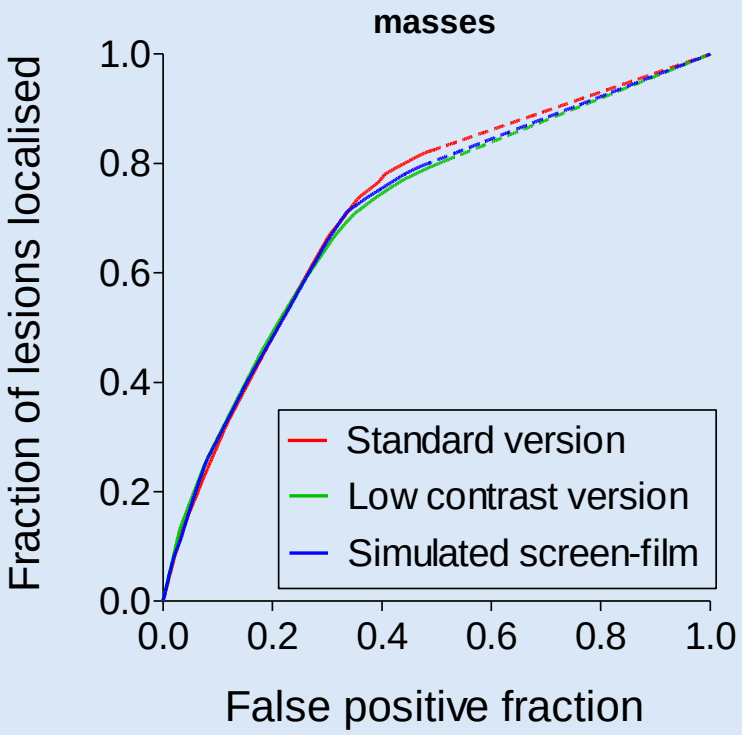




L Warren et al, AJR 2014

NCCPM

Image processing comparison



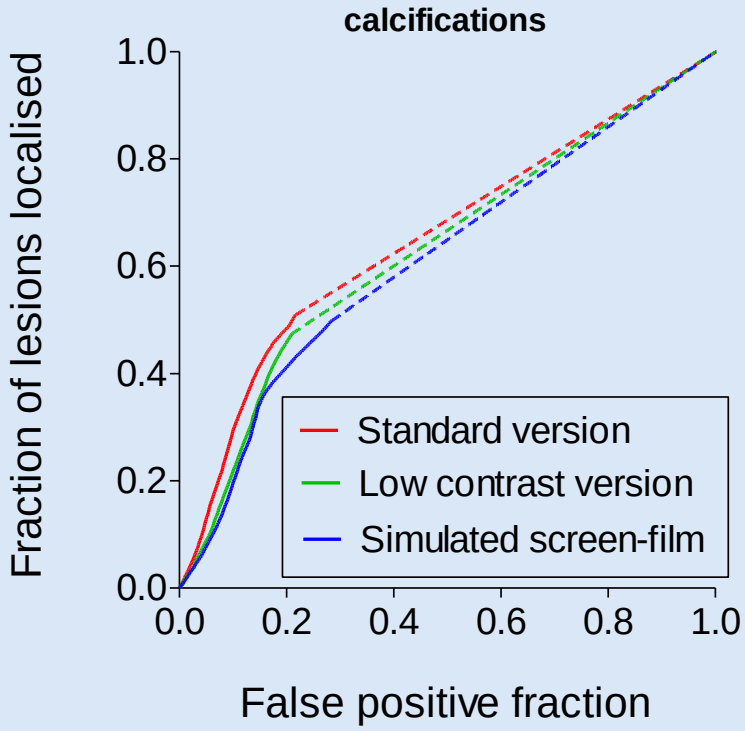


Image processing	Figure of merit	
	Masses	Calcifications
Standard	0.73	0.65
Low contrast	0.72	0.63 (p = 0.04)
Simulated screen-film	0.72	0.61 (p = 0.0005)

L Warren et al, AJR 2014

NCCPM

Conclusions from Observer Study 2

1. For the detection of calcification clusters the standard image processing was significantly better than the low contrast or simulated screen-film processing
2. For the detection of non-calcification lesions there was no significant difference was found between any of the three image processing pairs
3. There were no significant differences in the number of false positive non-calcification marks for any of the image-processing pairs
4. The number of false positive calcification marks increased by 15% for film-screen processing compared to standard processing

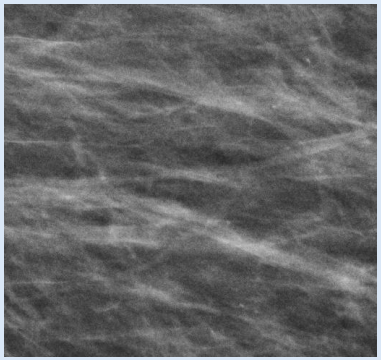
OBSERVER STUDY 3

Effect of detector type on detection of calcification and non-calcification lesions

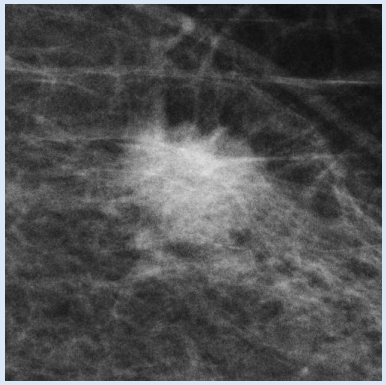
Image set for study

270 unprocessed image pairs – both breasts, one view (either CC or MLO)

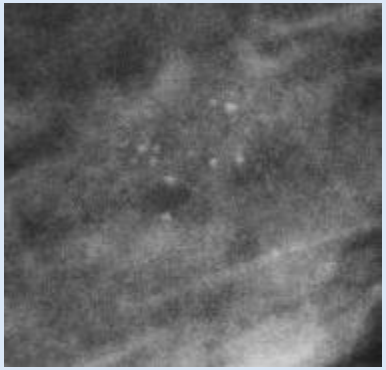
80 with no cancer present



80 containing malignant subtle non-calcified cancers



80 with inserted malignant subtle calcification clusters



30 containing biopsy proven benign lesions

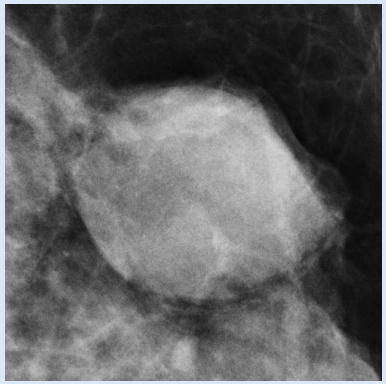


Image conversion

- **Starting image set was converted to:**
  - Arm 1: ‘a-Se’ detector
  - Arm 2: ‘CsI’ detector
  - Arm 3: ‘NIP CR’ detector
  - Arm 4: ‘Powder phosphor CR’ detector

  All doses reduced by 20%

  AGD = 1.08 mGy for 50 to 60 mm breast thickness

- **Image processing**
  - Agfa Musica[2]
  - Suitable for a wide range of image qualities

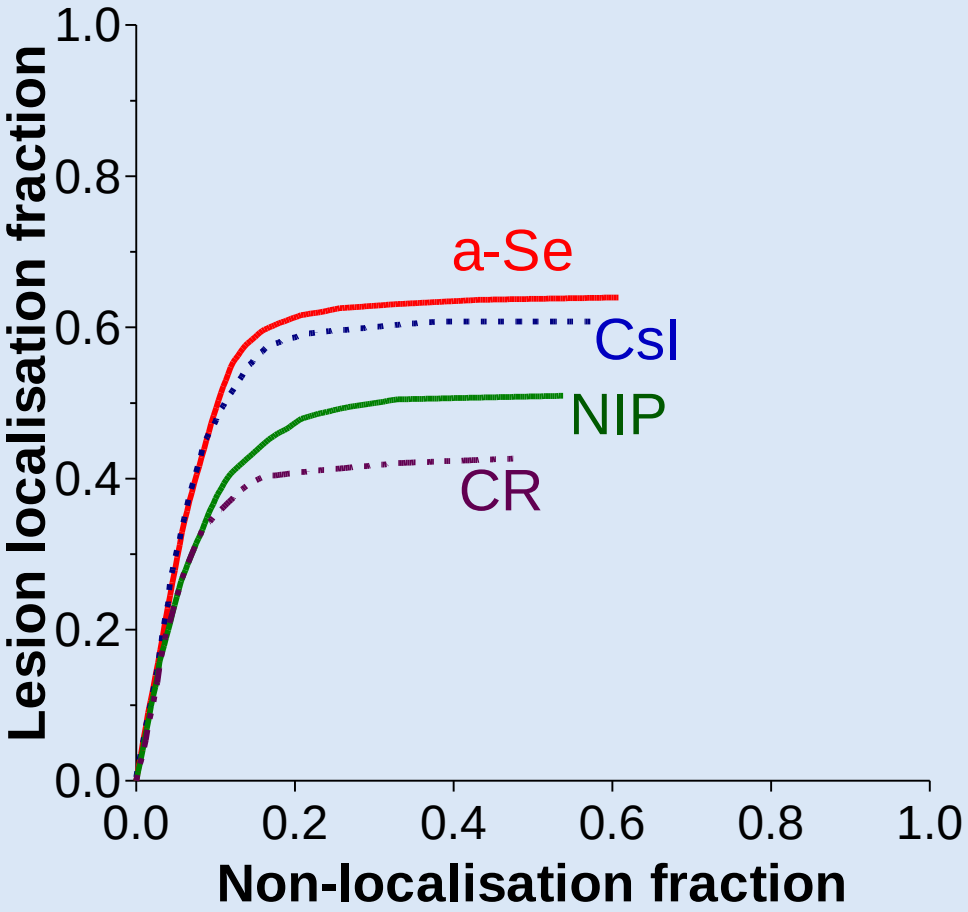
Calcification cluster detection and recall

Would you recall on the basis of this lesion?

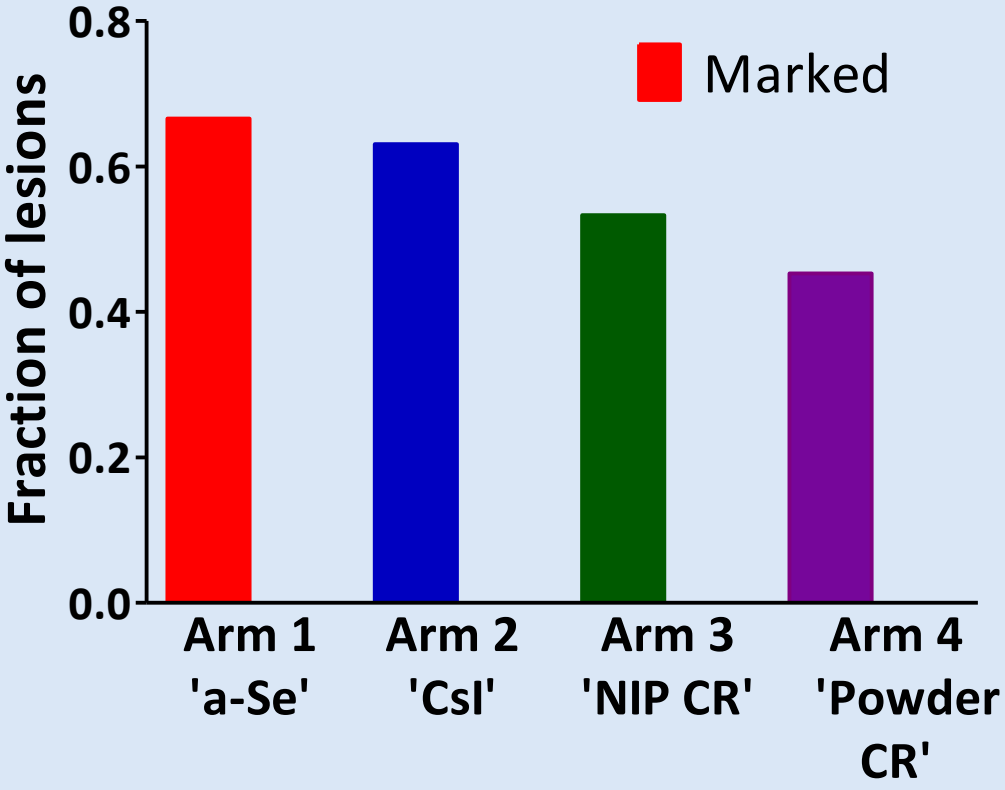
No: Very confident
No: Moderately confident
No: Slightly confident
Yes: Slightly confident
Yes: Moderately confident
Yes: Very confident

Calcification cluster detection and recall

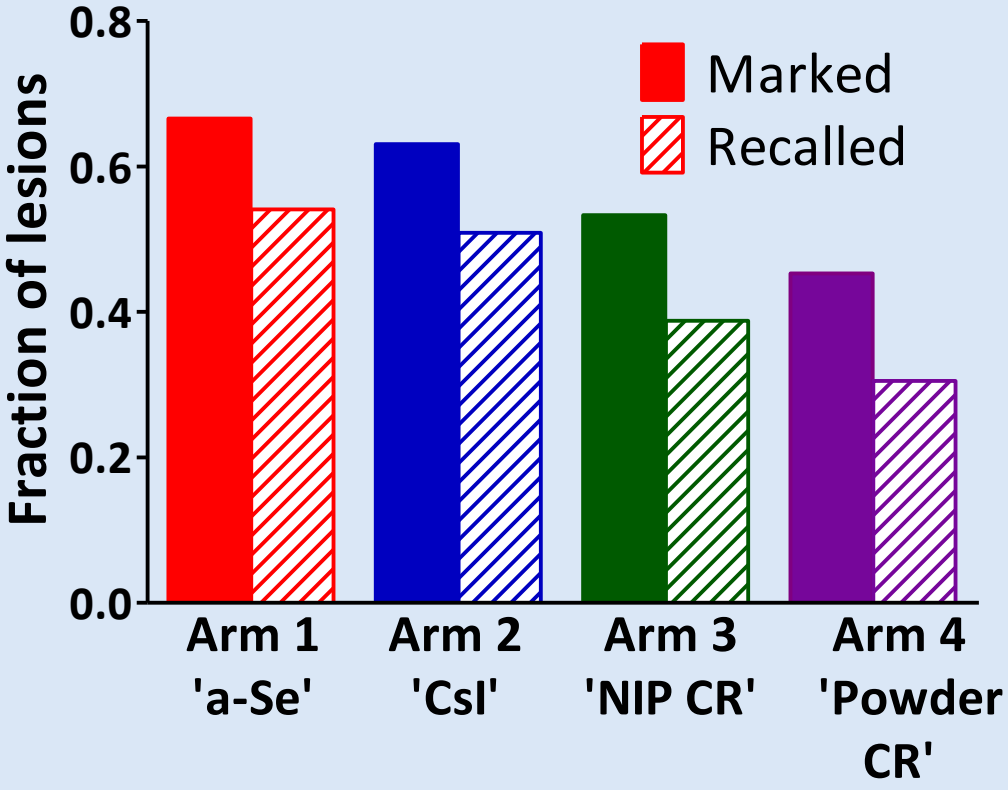
IQ pairs	*p*-value
a-Se v. CsI	**0.27**
a-Se v. NIP	**<0.0001**
a-Se v. CR	**<0.0001**
CsI v. NIP	**0.0001**
CsI v. CR	**<0.0001**
NIP v. CR	**0.048**



Calcification cluster detection and recall



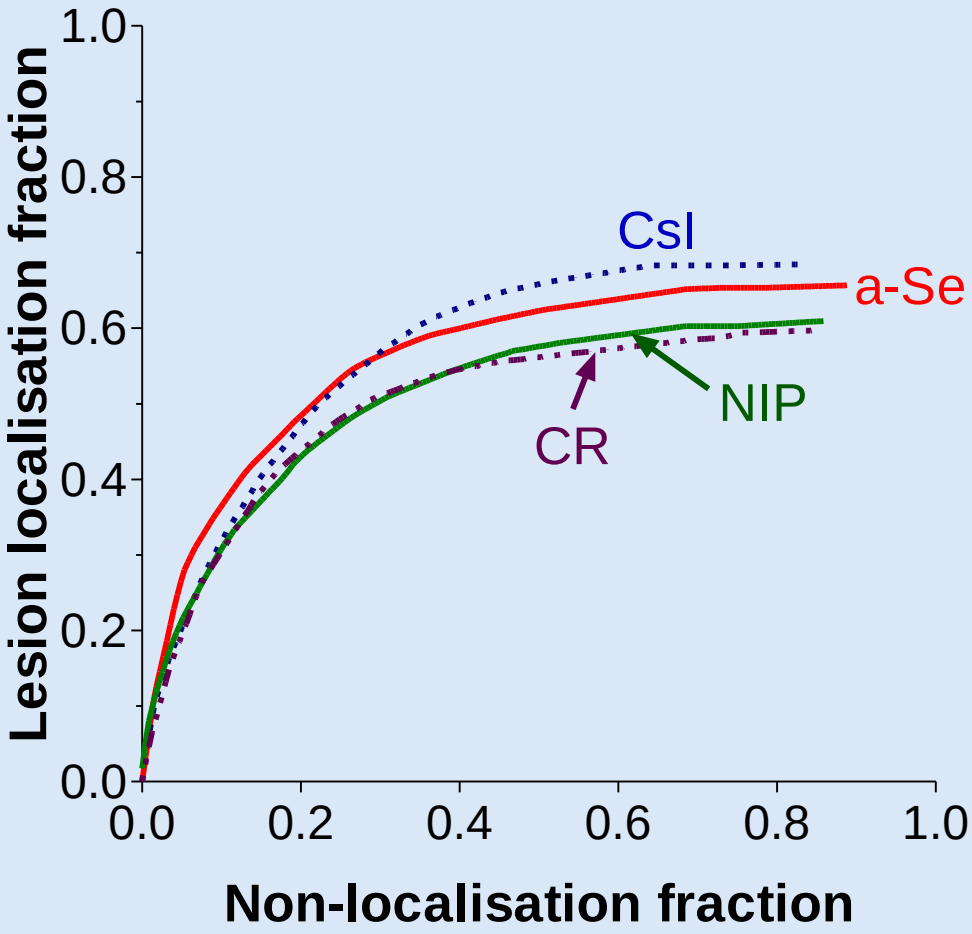
Calcification cluster detection and recall



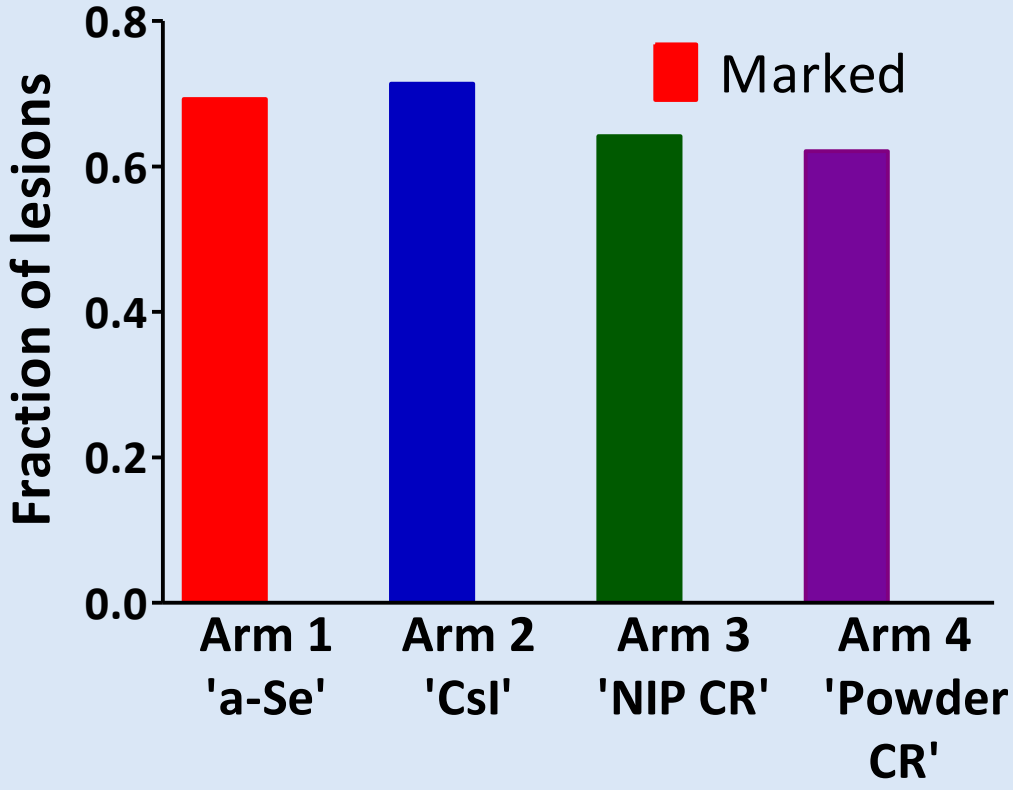
Compare with **a-Se**	*Change in recall*
CsI	**n.s.**
NIP	**-28%**
CR	**-44%**

Non-Calcification detection and recall

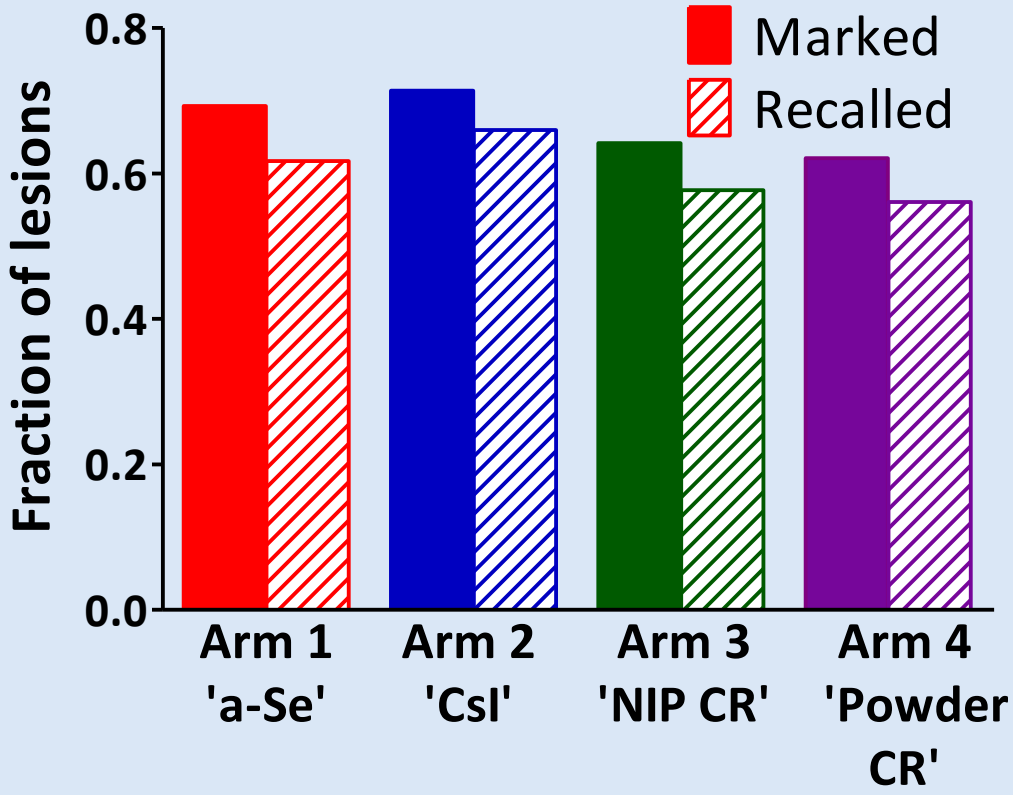
IQ pairs	*p*-value
a-Se v. CsI	**0.93**
a-Se v. NIP	**0.002**
a-Se v. CR	**0.0007**
CsI v. NIP	**0.002**
CsI v. CR	**0.0009**
NIP v. CR	**0.77**



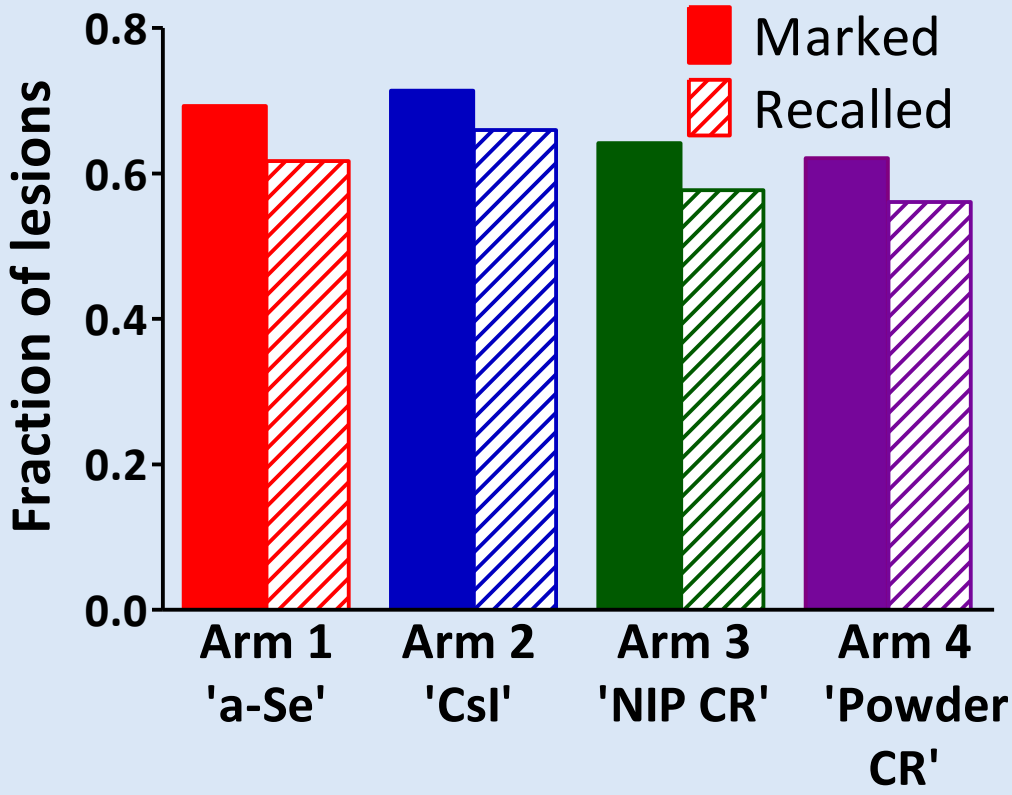
Non-Calcification detection and recall



Non-Calcification detection and recall



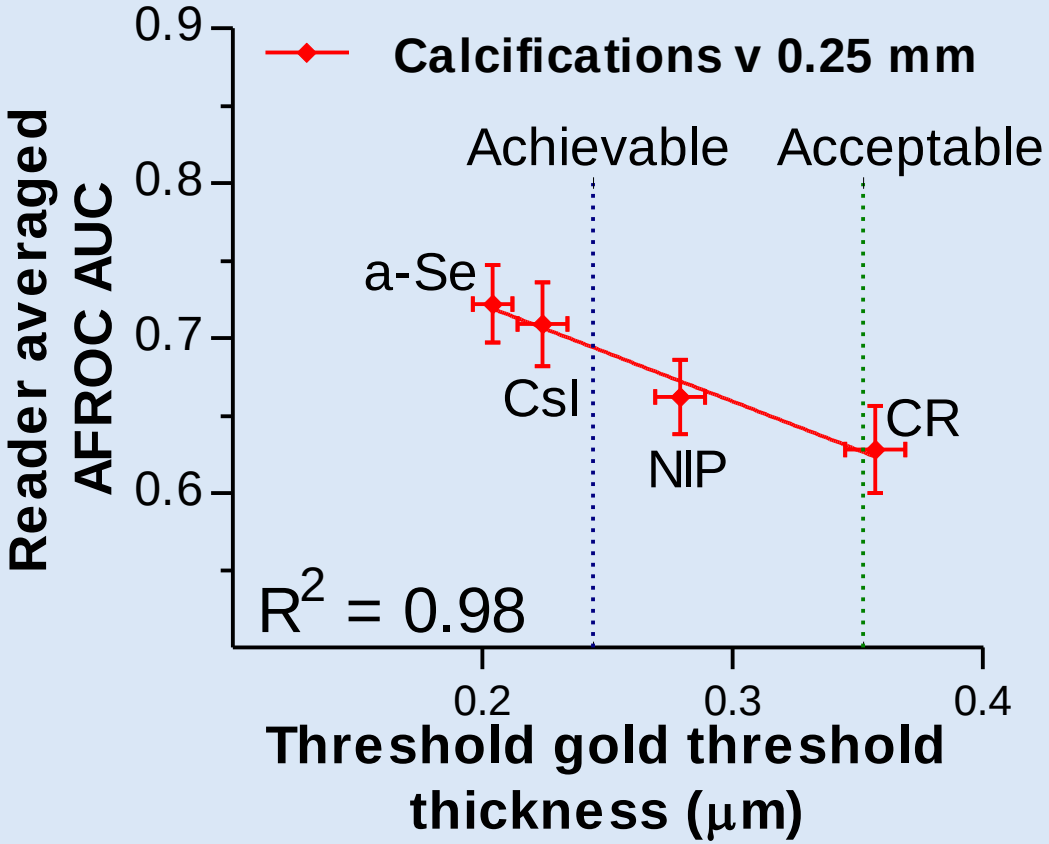
Non-Calcification detection and recall



Compare with **a-Se**	*Change in recall*
CsI	**n.s.**
NIP	**-6.5%**
CR	**-9.1%**

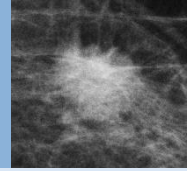
Relationship between physical image quality & cancer detection?

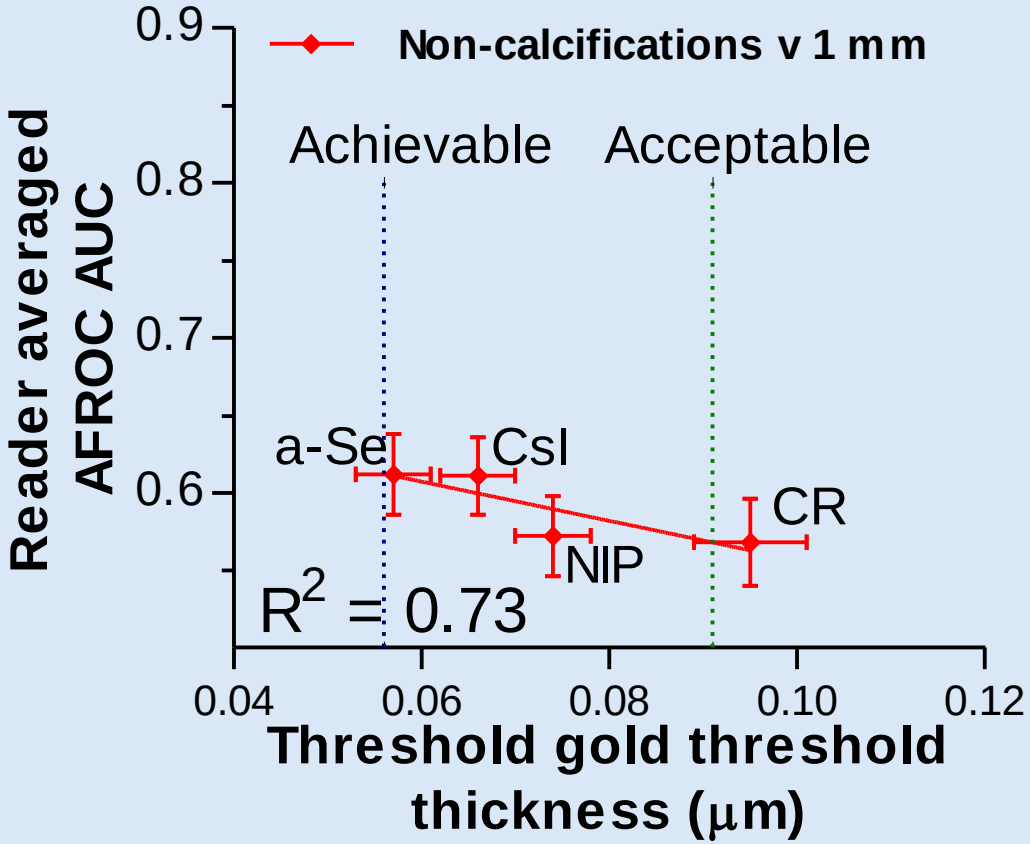
Calcification clusters



Clear relationship between calcification detection and CDMAM results

Non-calcification lesions





Relationship is not so clear

Larger detail
More complex task

Conclusions: 3 OPTIMAM observer studies

1. Threshold gold thickness using the CDMAM phantom is a good predictor of detection of clinical calcification
2. Traditional CR systems poorer than DR for calcification detection even at relatively high dose levels. NIP may be acceptable at high dose levels
3. EU guidelines for IQ are clinically relevant
4. Good IQ is important for calcification detection and to a lesser extent for masses.

Conclusions: 3 OPTIMAM observer studies

1. Systems should exceed achievable levels in EU guidelines.
2. Acceptable level is too low. Guidelines may need revision
3. Image processing has an effect on calcification detection

Thank you for your attention