Royal United Hospitals Bath
NHS Foundation Trust

# Alternate Image Quality Metrics for Advanced Reconstruction Algorithms

Laurence King, DClinSci
Principal Physicist
Royal United Hospitals Bath

The RUH, where you matter

# motivation

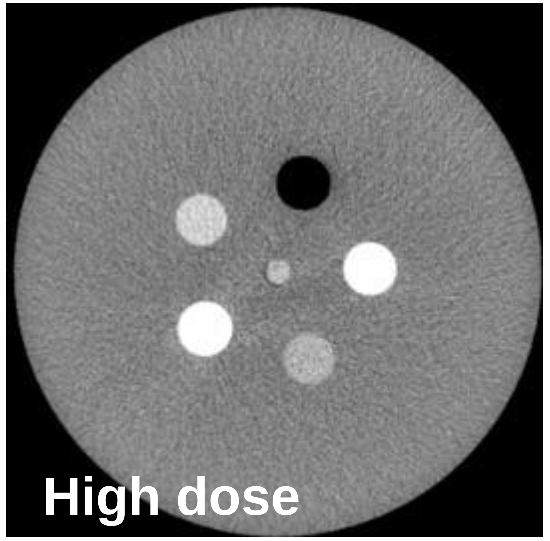
DClinSci research project carried out as part of HSST training

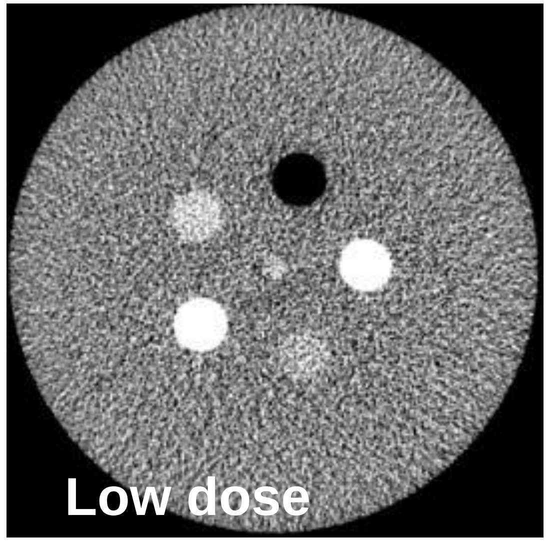
# Filtered back projection (FBP)

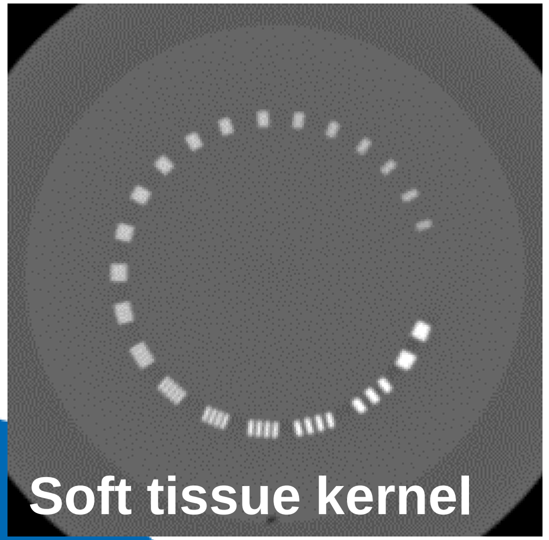
- Explicit, analytic mathematical solution to CT reconstruction
- For an object of a known attenuation there is a well-understood link between image noise and acquisition dose

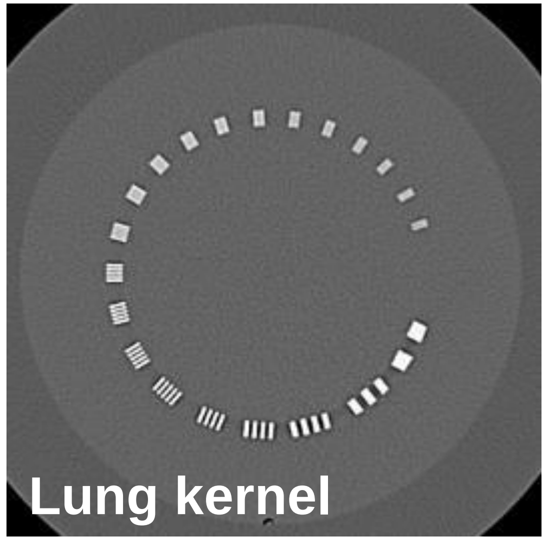
- The convolution kernel applied to data before back projection determines the trade off between image noise and spatial resolution
- For a given convolution kernel we are able to predict image noise and spatial resolution properties for a given object

- Image quality measured in a phantom is a viable predictor of clinical image quality.



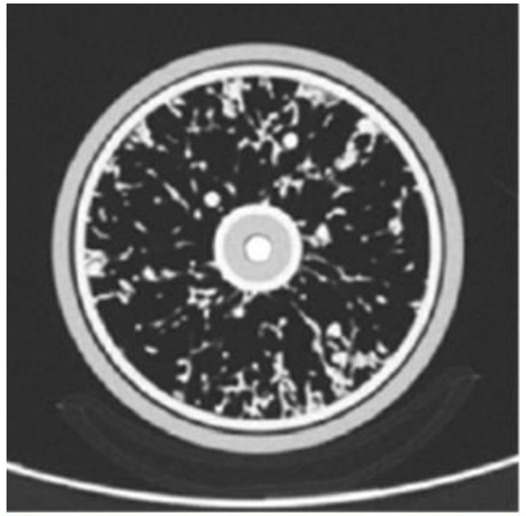
# Iterative reconstruction (IR)

- Proprietary black boxes
- Often starts as FBP, followed by iterative process of forward-projection and comparison of simulated raw data to actual acquisition data and subsequent correction of reconstructed image set
- Options may include:
  - Preferential weighting of less noisy projection data
  - Modelling of photon optics and scatter properties (full Model-Based IR)
  - Additional noise reduction by identification of statistical noise (in raw data and in image space)
- The outcome is highly selective noise reduction:
  - Uniform areas experience high levels of noise reduction
  - Areas containing structures experience lower noise reduction in order to preserve edge appearance

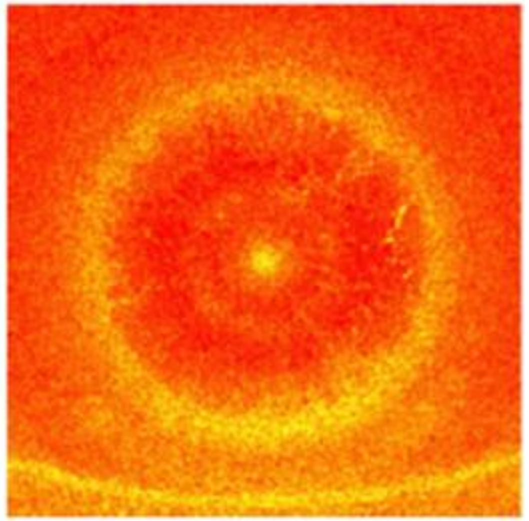
# Challenges of IQA with IR

Solomon et al, Med. Phys. 41 (9), September 2014
http://dx.doi.org/10.1118/1.4893497

Phantom image – realistic lung parenchyma structures

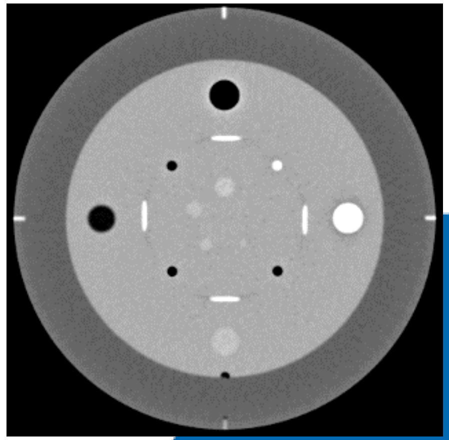


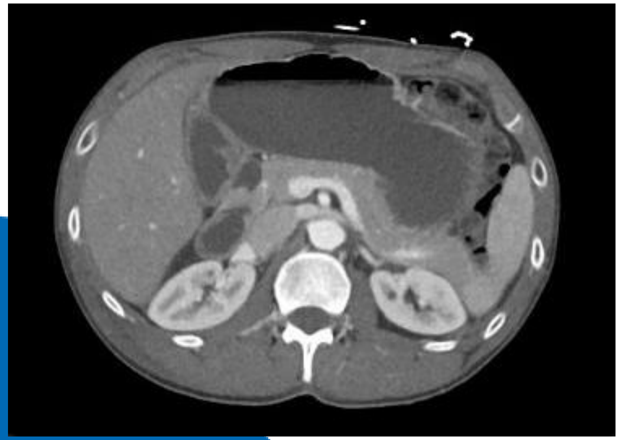
Filtered back projection noise magnitude map

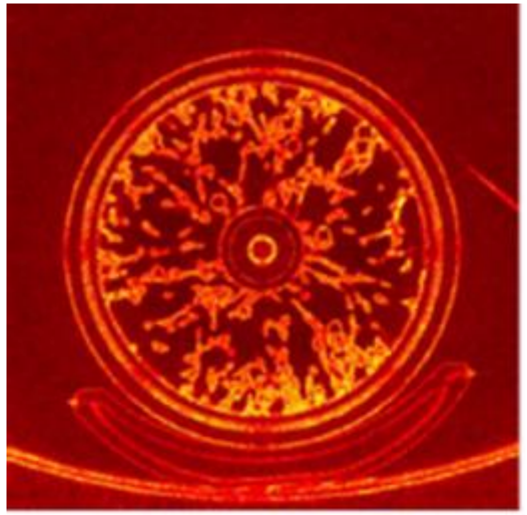


- The selective noise reduction properties of IR means that noise reduction becomes locally dependent on object spatial frequency content and contrast.
- Reconstructed Image quality is dependent on the object you are scanning.

- Traditionally we measure image quality in mostly uniform phantoms with few different contrast inserts
- Traditionally we assess image resolution only at high contrasts
    - <u>This is no longer adequate for IR images.</u>

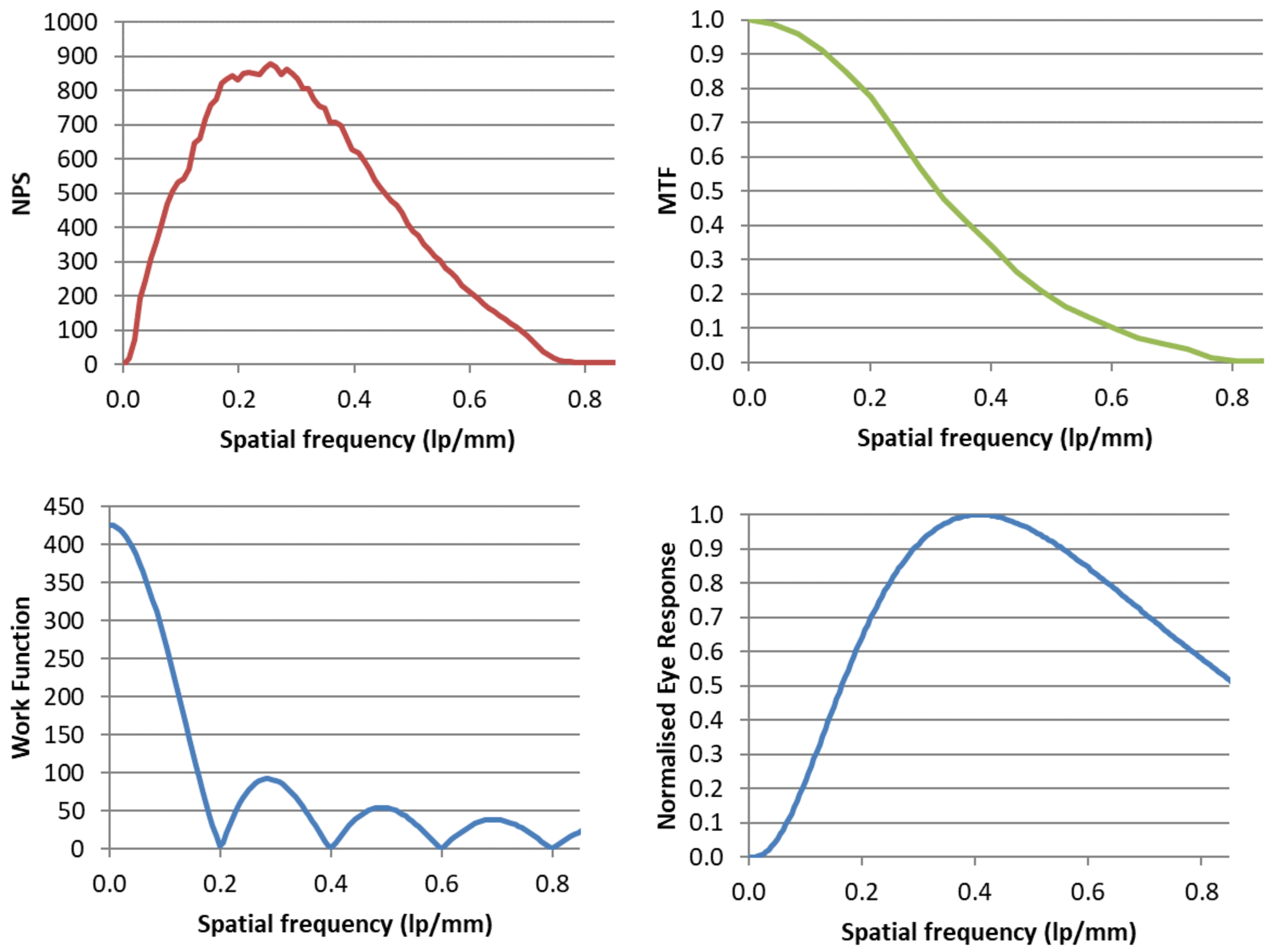






# AAPM report 233 approach to IQA

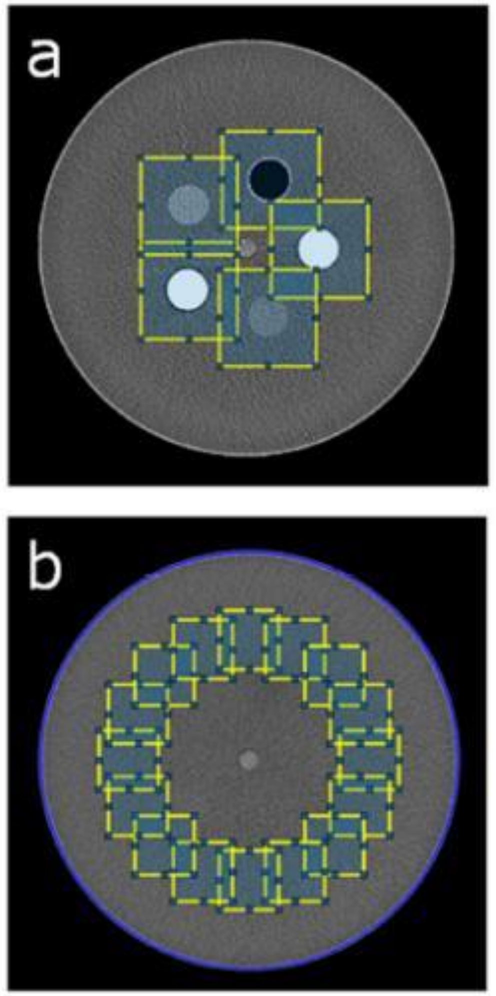
- AAPM report 233 recommends measurement of task-specific image quality metrics:
    - NPS in specified object at a specified dose level
    - TTF = MTF for a specified object contrast in the same size phantom and dose
- A detectability index d’ can then be calculated for an object of that contrast in a noisy background:




$$d'^2_{NPWE} = \frac{\left[\int_0^{f_N} |W(f)|^2 \cdot TTF^2(f) \cdot E^2(f)\, df\right]^2}{\int_0^{f_N} |W(f)|^2 \cdot TTF^2(f) \cdot NPS(f) \cdot E^4(f)\, df}$$

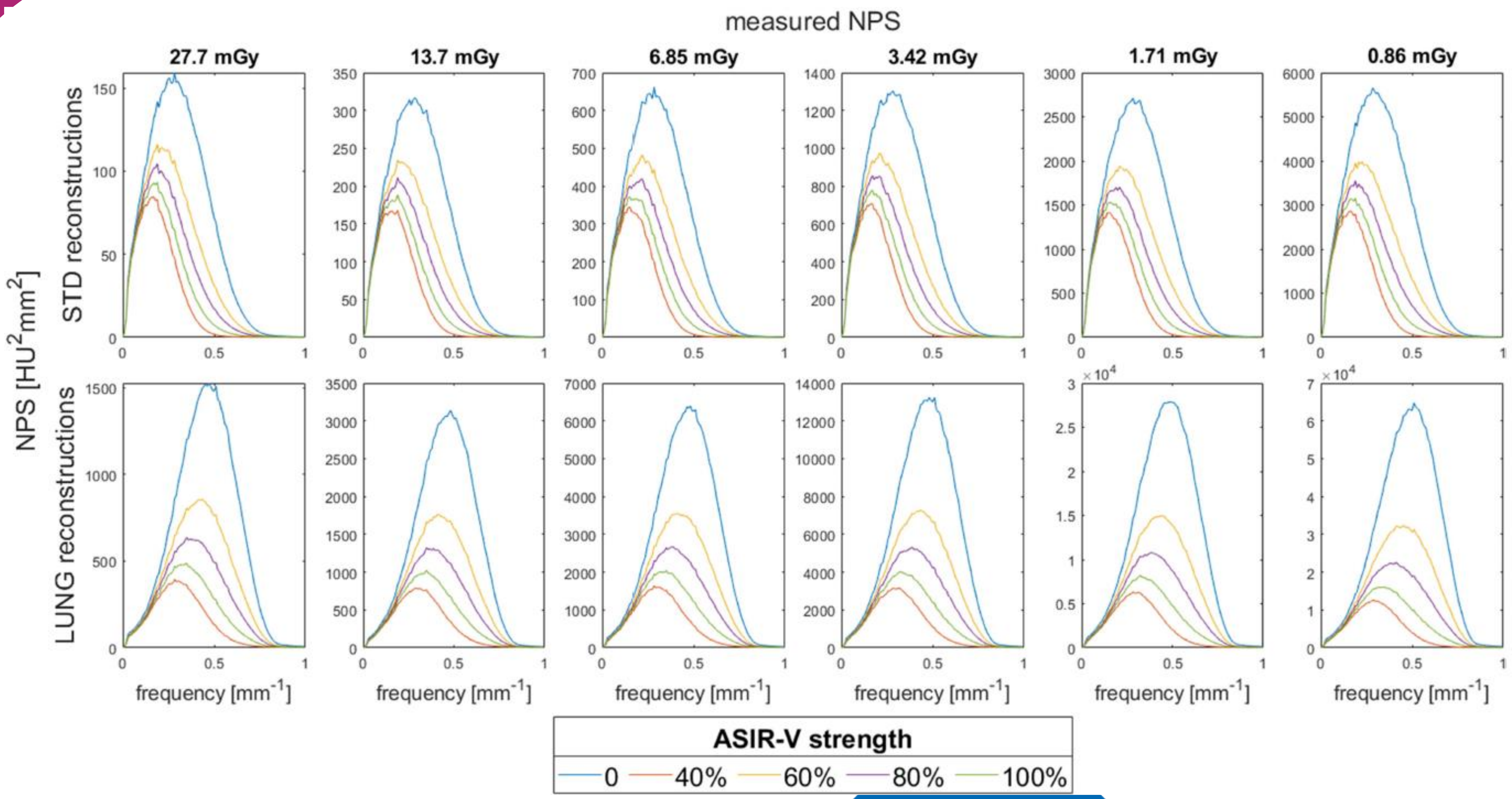
# AAPM report 233 approach to IQA

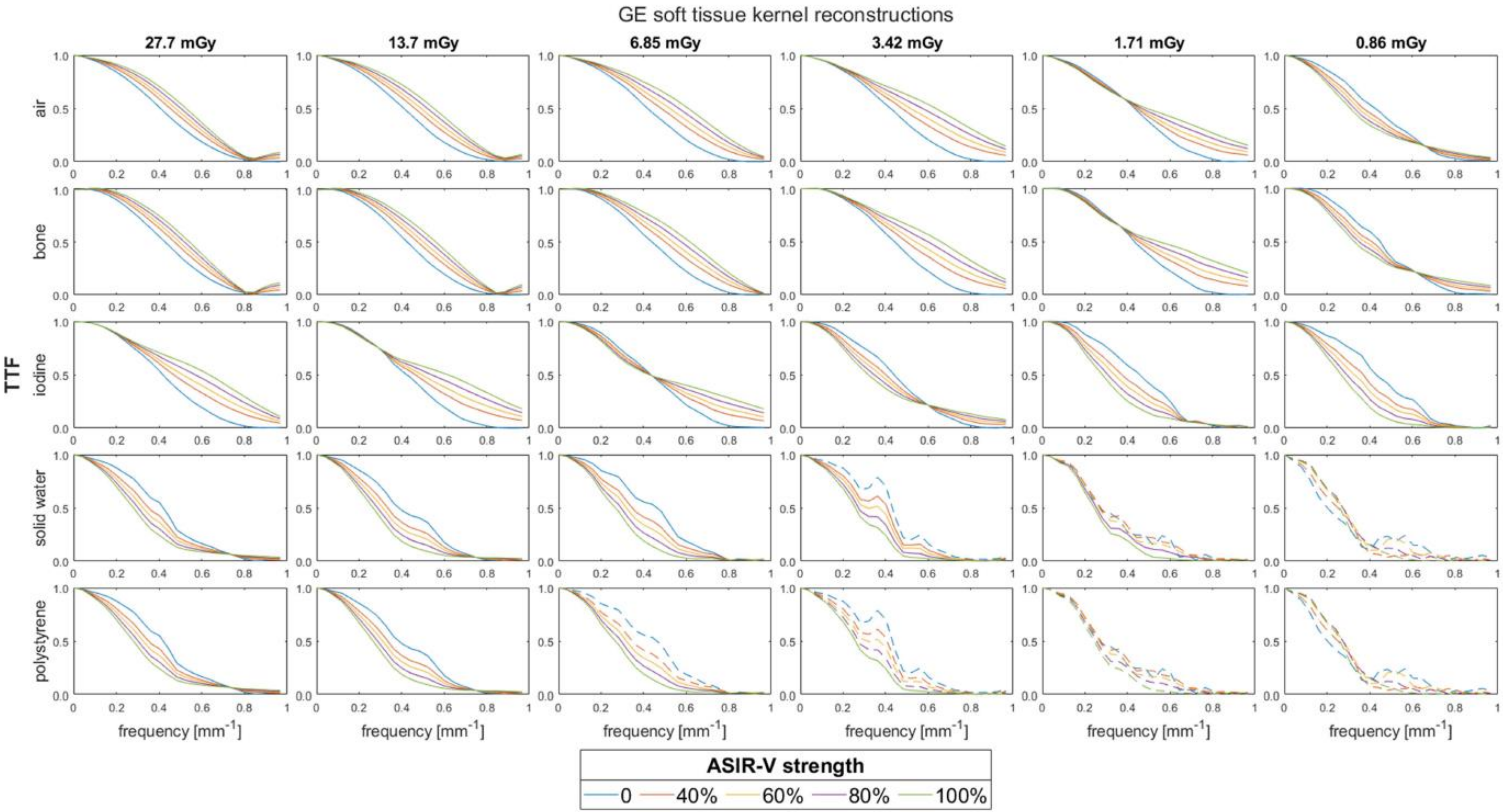
- This requires many measurements at different dose levels and contrasts…




| Insert material | HU contrast with polyethylene background |
|---|---|
| Air | -895 |
| Bone (50% CaCO3) | 1000 |
| 10 mg/mL iodine | 335 |
| 1451 CT HE solid water | 90 |
| Polystyrene | 50 |

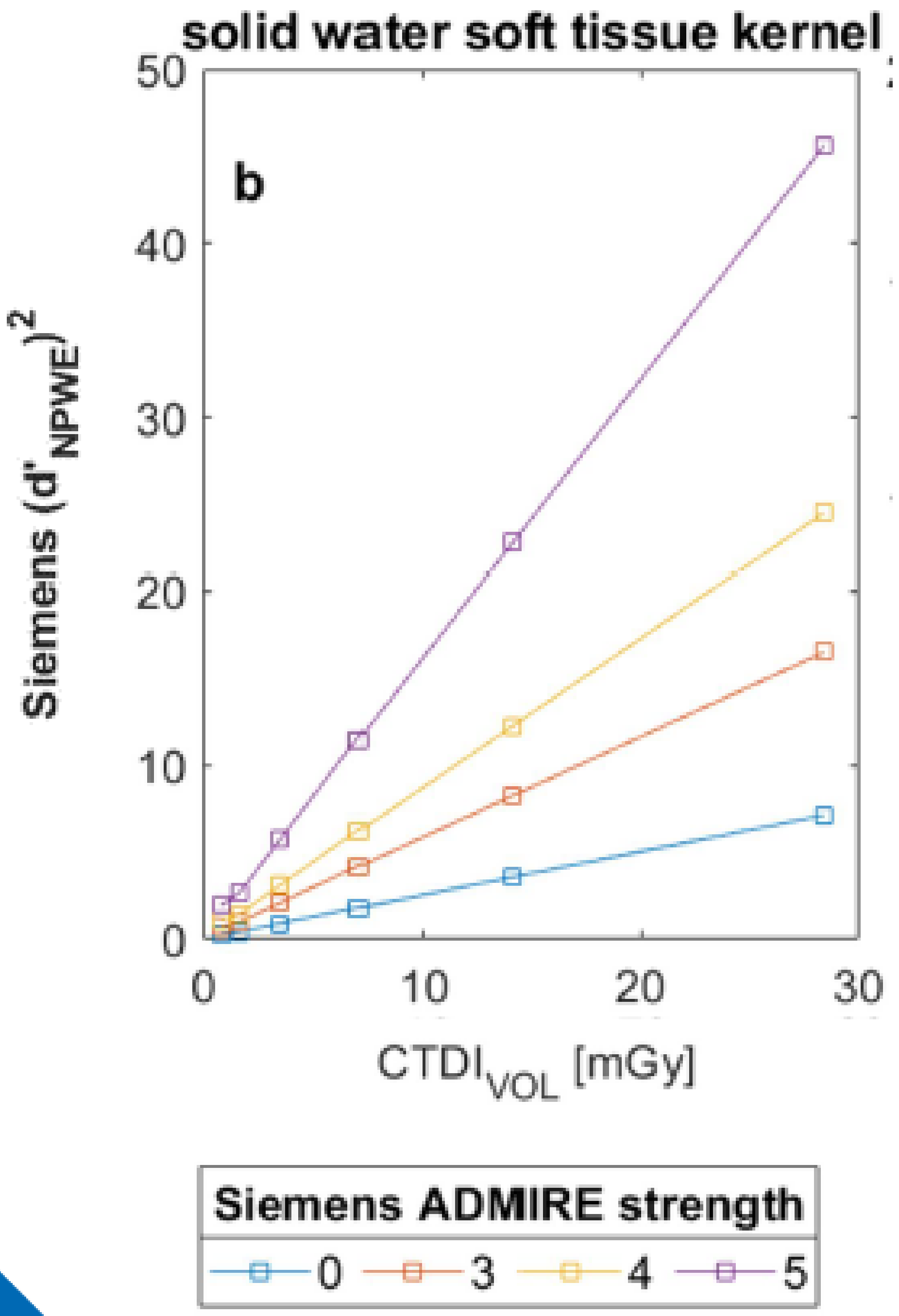
Th


measured NPS
27.7 mGy
13.7 mGy
6.85 mGy
3.42 mGy
1.71 mGy
0.86 mGy
NPS [HU$^2$mm$^2$]
STD reconstructions
LUNG reconstructions
frequency [mm$^{-1}$]
ASIR-V strength
0
40%
60%
80%
100%


GE soft tissue kernel reconstructions
27.7 mGy
13.7 mGy
6.85 mGy
3.42 mGy
1.71 mGy
0.86 mGy
TTF
air
bone
iodine
solid water
polystyrene
frequency [mm-1]
ASIR-V strength
0
40%
60%
80%
100%

# AAPM report 233 approach to IQA

- Results: plots of contrast detectability index against acquisition dose for specific contrast object, object size and reconstruction setting.
- However:
    - measurements are still made in mainly uniform objects with no realistic tissue structure or background.
    - We know that this will effect contrast detection and spatial resolution for IR algorithms.
    - Not all clinical tasks are based on contrast detection – so is calculation of contrast detectability always clinically relevant?



# Seeking a practical image quality metric for clinical optimisation

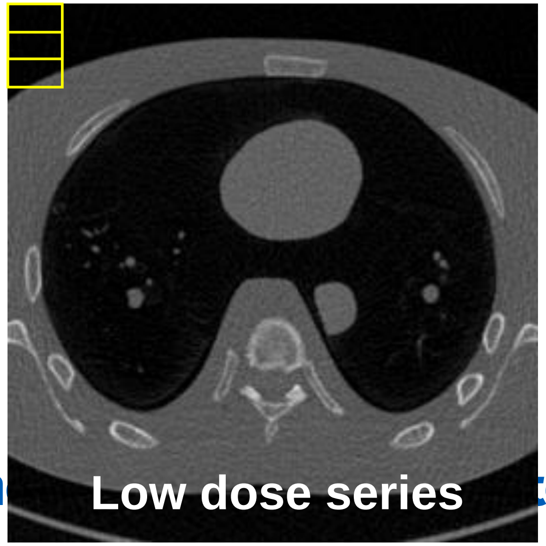
# Investigating other approaches - SSIM

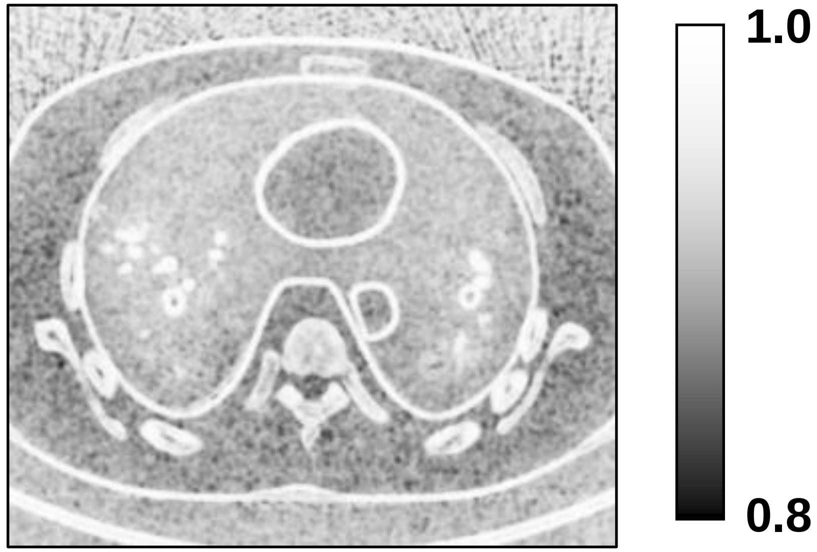
- The Structural Similarity metric (SSIM) has been used extensively in imaging research for over a decade
  - SSIM is a full-reference image quality metric
  - You need a “ground truth” reference image
  - You then assess image quality of test images against ground truth images
  - SSIM is calculated in a pixel-wise nature to give relative image quality between the image sets
  - SSIM is then pooled over your region of interest to give a single image quality metric for the image, between 0 and 1, where 1 = exact similarity.

# Calculating SSIM

Luminance · Contrast · Structure

$$SSIM(x,y) = \left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}\right) \cdot \left(\frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}\right) \cdot \left(\frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}\right)$$


SSIM map

In local pixel patches x and y:

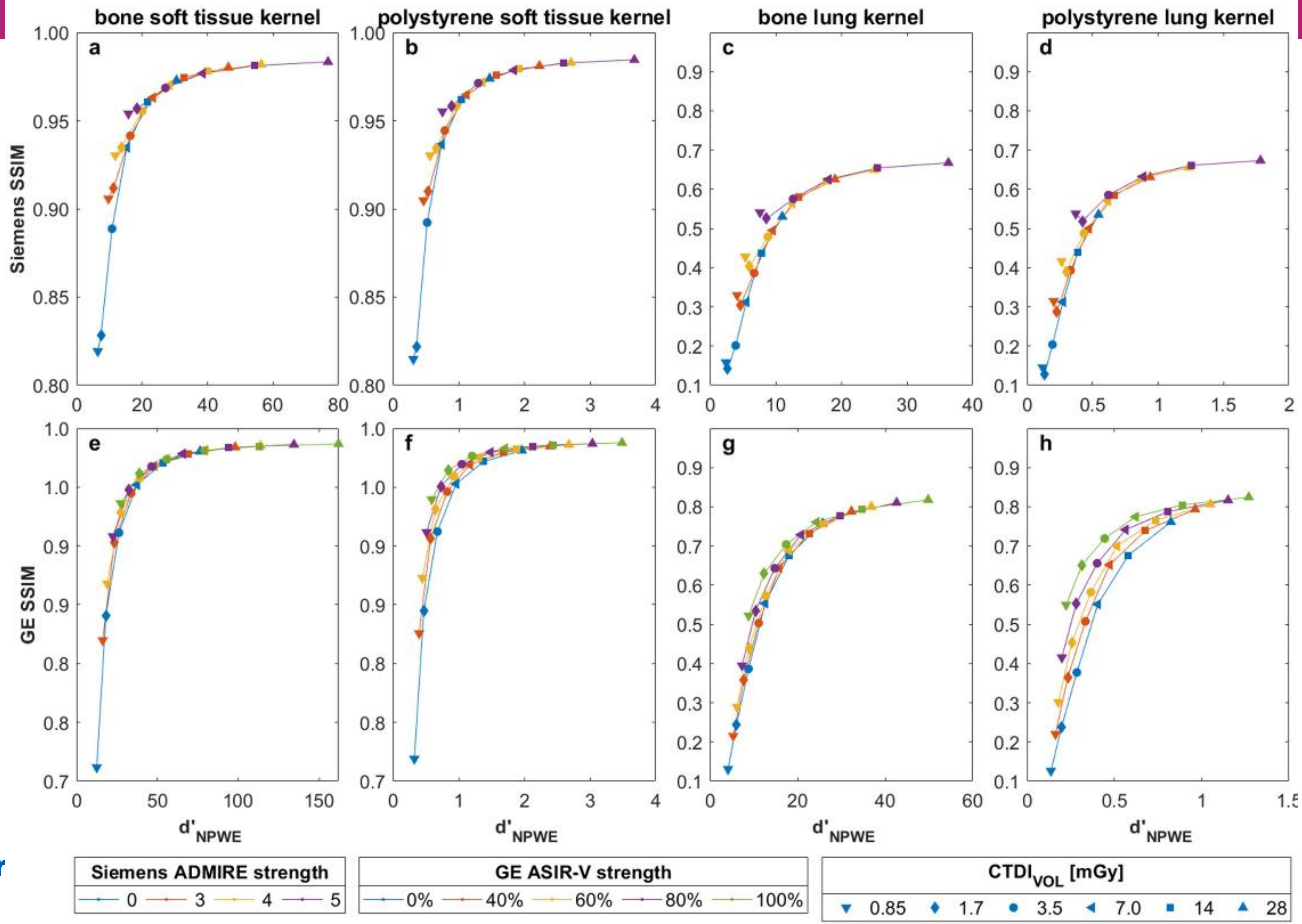
$\sigma$ = standard deviation

$\mu$ = mean

$\sigma_{xy}$ = pixel covariance

SSIM takes value between 0 & 1

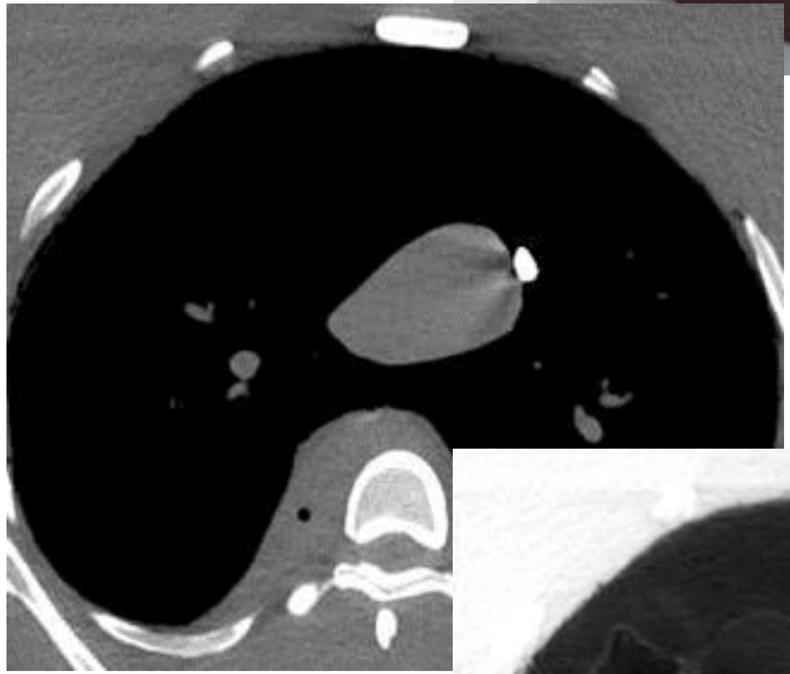
# Assessing SSIM in a traditional phantom




The RUH, wher

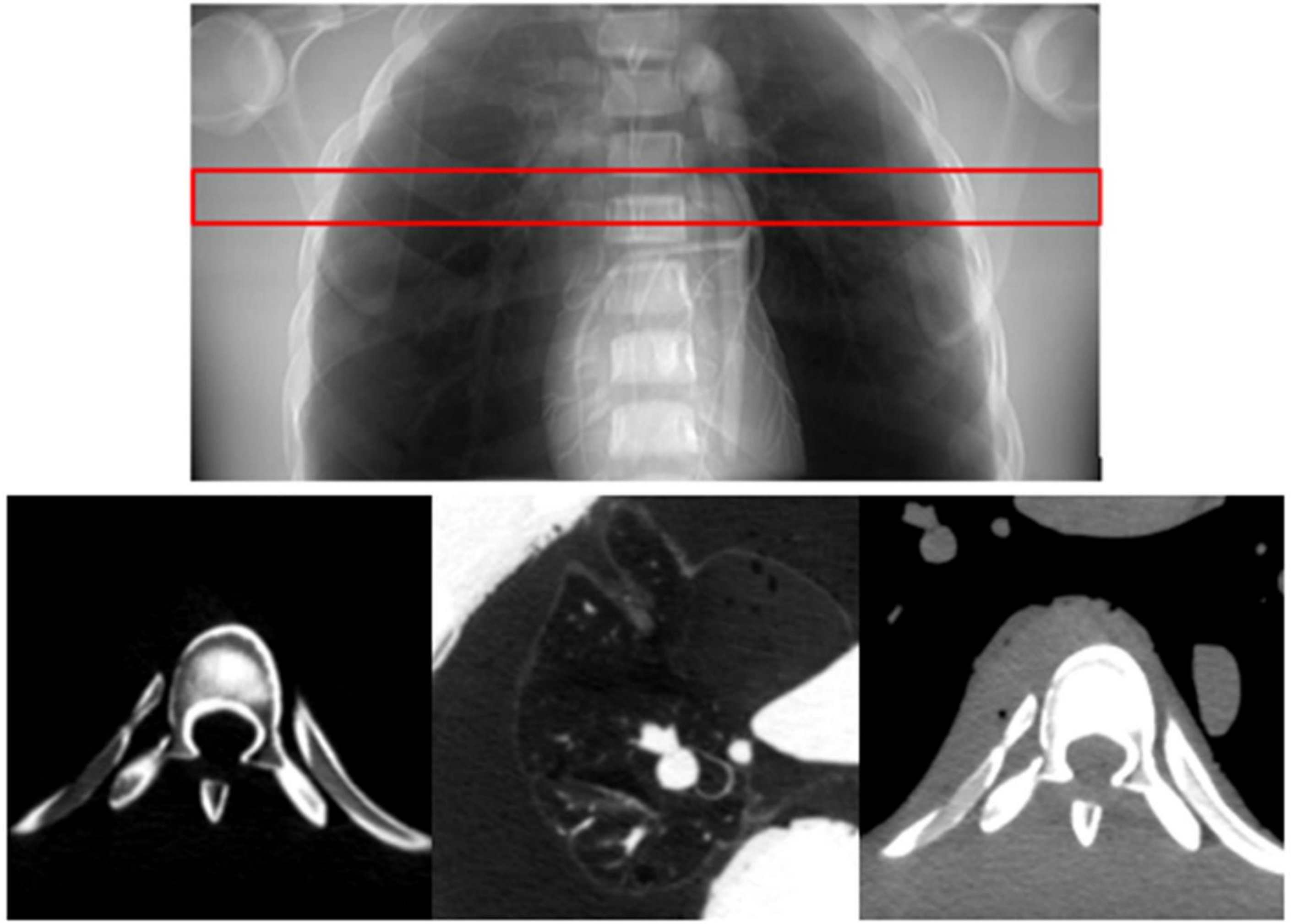
# Assessing SSIM in an anthropomorphic phantom

- The aim was to determine if SSIM could be used to predict image quality assessment by a radiologist
- SSIM was designed to reflect human visual response in detecting changes in luminance, contrast and structural content in an image
- SSIM can be calculated from images of ANY object – no need to stick to circular inserts in a uniform phantom.
- So… use a clinically relevant, task-specific object such as a realistic lung phantom

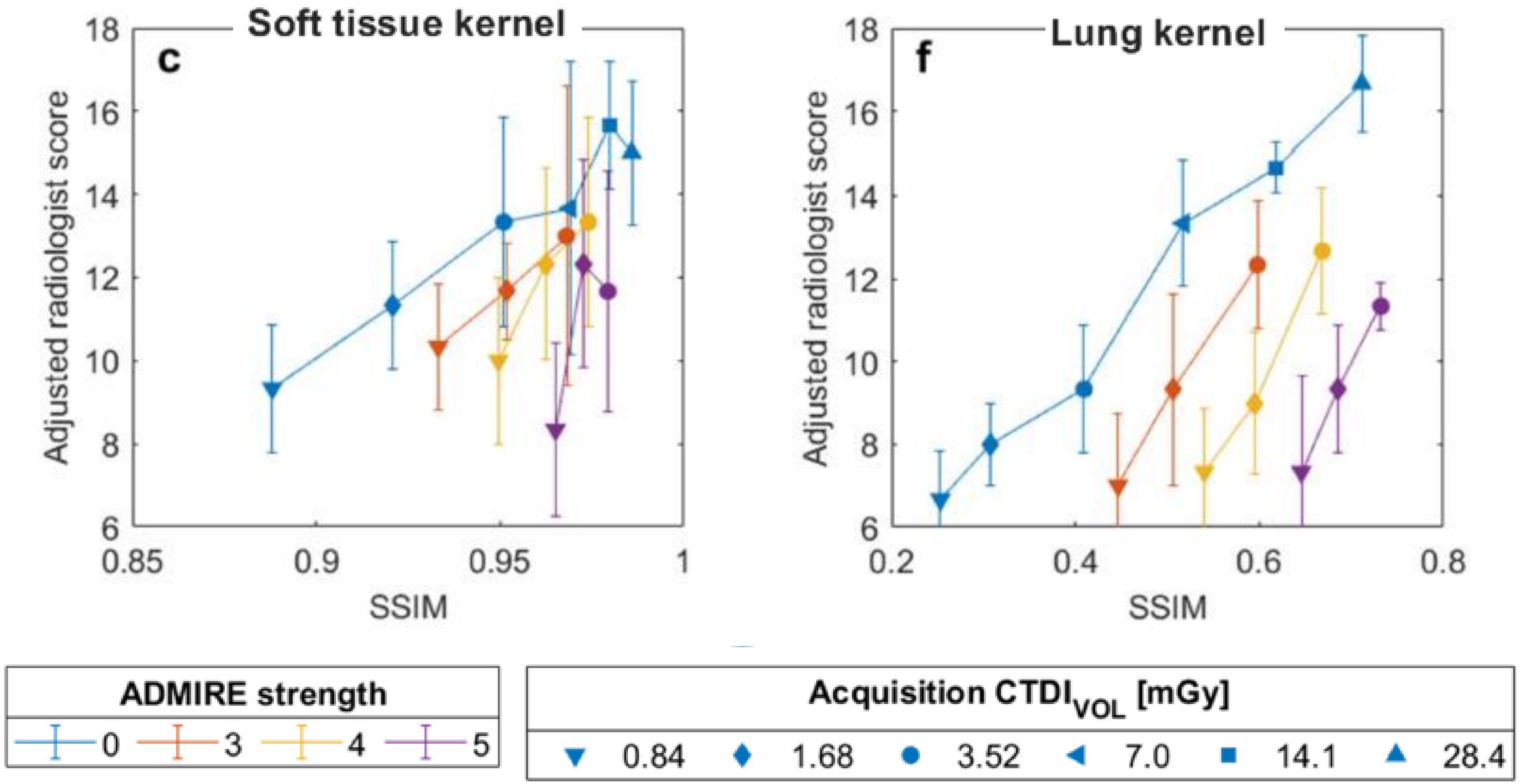


# Assessing SSIM in an anthropomorphic phantom

- Chest phantom scanned at many doses and IR reconstruction options
- Image quality criteria scored subjectively by three radiologists based on European Guidelines on Quality Criteria for Computed Tomography
- Six clinical structures scored on a 4-point Likert scale for each image set.
- SSIM calculated relative to high dose FBP image to represent “ground truth”



# Assessing SSIM in an anthropomorphic phantom



# Conclusions?

- Interesting but not definitive results. More research needed!
- SSIM *might* be an effective way to predict clinical image quality against a reference
    - SSIM is essentially a test of **fidelity of image reconstruction against a ground truth**
- SSIM is probably less sensitive to changes in noise texture than *d'*.
- Other studies use different ground truth images, e.g. the noise-free 3D print file for a printed phantom, or artificially de-noised images.
- SSIM was pooled over the entire phantom region – a next step would be to calculate tissue by tissue in the phantom
    - Alternative SSIM-derived metrics exist that perform weighted pooling of SSIM by classifying the images into uniform, textured and edge structures.

# Take home messages

- We ought to think more deeply about how, and why we are assessing image quality in CT

- Quality Control: technical or physical assessment of CT scanner functionality
  - Requires objective and reproducible measurement of image quality metrics
  - Ensures that we detect faults in physical imaging chain, changes in software

- Image optimisation and clinical performance
  - Our IQA needs to be more clinically relevant and task-based.
  - We should be performing IQA using clinically relevant test objects containing realistic anatomical detail and structure.
  - The search is still on for more relevant measures of clinical image quality.

# Image quality assessment

**Technical image quality** → **Clinical efficacy**

| Simple phantoms | | Anthropomorphic phantoms | Patient-specific |
|---|---|---|---|
| Generic tests, Performance specification | System characterisation, Protocol comparison | Task-specific tests, clinical detail-based performance | Diagnostic sensitivity and specificity |

# Thanks to

- Elly Castellano
- Marianne Aznar and Abigail Bryce-Atkinson
- My colleagues at the Royal United Hospitals Bath Medical Physics department

# References


- Performance Evaluation of Computed Tomography Systems - The Report of AAPM Task Group 233 (2019) https://www.aapm.org/pubs/reports/detail.asp?docid=186

- Zhou Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, (2004). ‘Image quality assessment: from error visibility to structural similarity,’ in IEEE Transactions on Image Processing, **13**(4), pp. 600-612 https://doi.org/10.1109/TIP.2003.819861

- Joemai RMS, Geleijns J. (2017). ‘Assessment of structural similarity in CT using filtered backprojection and iterative reconstruction: a phantom study with 3D printed lung vessels.’ *BJR* **90**(1079):20160519. https://doi.org/10.1259/bjr.20160519