Metrics for Volume CT: Characterizing Sources of Variation

Nicholas Petrick, Marios A. Gavrielides, Qin Li, Rongping Zeng, Berkman Sahiner, Kyle J. Myers

Center for Devices and Radiological Health, U.S. Food and Drug Administration

Purpose

- Discuss various studies and approaches for characterizing sources of variation in CT volumetry
  - Concentrate on how phantom studies can systematically probe, identify and potentially minimize measurement error
- Updated version of RSNA 2012 Refresher Course

Outline

- Background
- Part 1
  - Phantom and synthetic lesion designs
- Part 2
  - Performance metrics
- Part 3
  - Lessons from the literature, QIBA, FDA studies
- Summary

Background

- Use of imaging biomarkers is limited by estimation uncertainty

Factors influencing size measurement accuracy

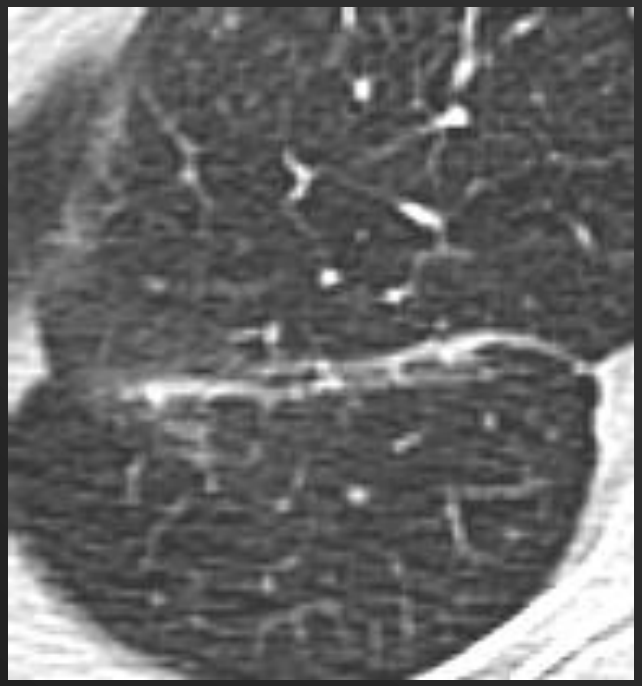
- Patient factors
  - Natural variation in nodules/patients
- CT acquisition parameters
  - Slice thickness, collimation, exposure, etc
- CT Reconstruction parameters
  - Reconstruction type, filter, etc
- Measurement tool effects
  - Size estimation method, thresholds, seed points, readers, etc
- Nodule characteristics
  - Size, shape, density, location and background, vascular/pleura attachments, etc

Questions to address

- Which parameters substantially contribute to measurement error?
- How can they be controlled to minimize error?

Approaches to assess estimation performance

- Analysis of clinical data
  - True size/change in size generally unknown
  - Limited to variance/ agreement analyses



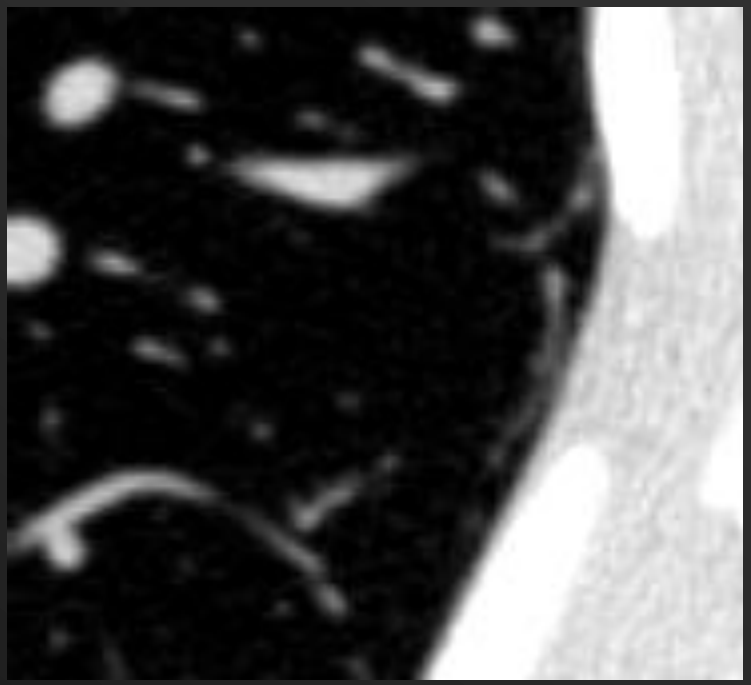
RIDER MSK-2283289298

Approaches to assess estimation performance

- Analysis of synthetic data
  - In silico data
    - Simulated image sets
    - Truth known
  - Physical phantom data

Why physical phantoms data for CT

- Fully incorporate scanner/acquisition physics
- Truth generally known
  - Be it lesion size or change in size
- Easy to collect scans across imaging, lesion characteristics
  - No patient exposure
- Analysis
  - Bias and variance analysis
  - May provide bound on estimation performance



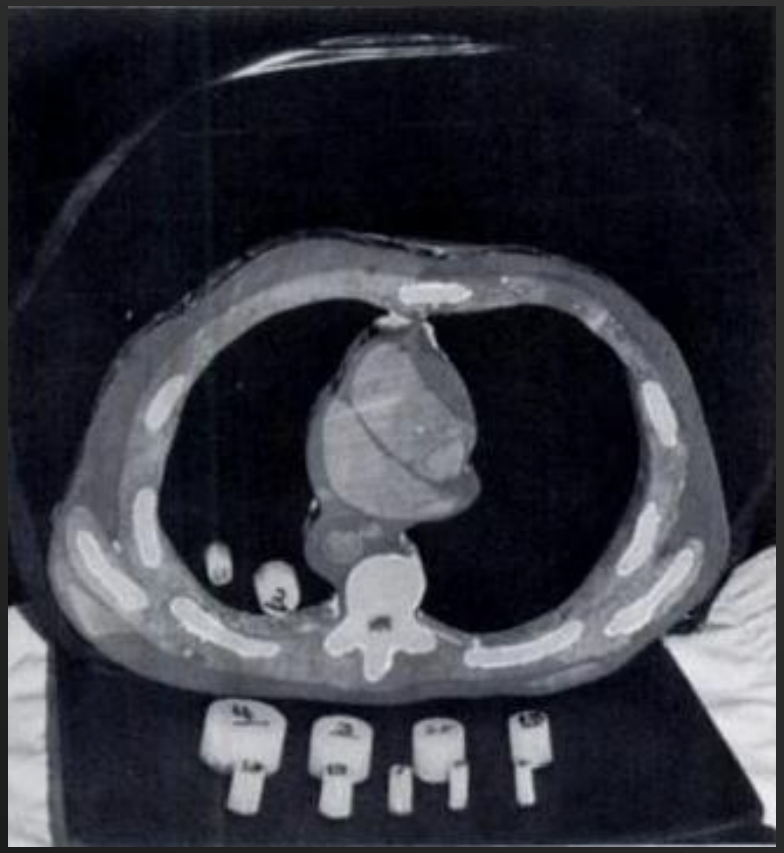
FDA-018-6591

Factors assessable with phantom data

- ✗ Patient factors
- ✓ CT acquisition parameters
- ✓ CT Reconstruction parameters
- ✓ Measurement tool effects
- ? Nodule characteristics

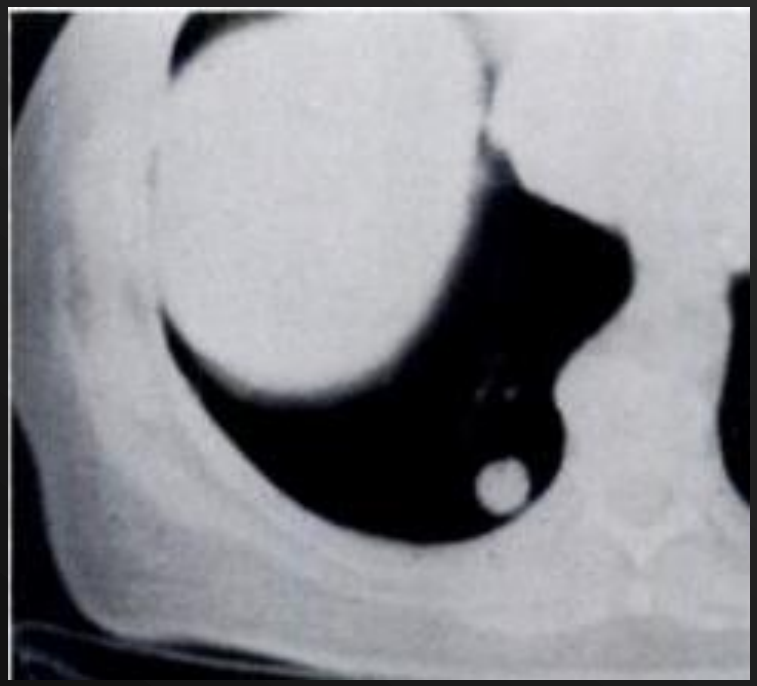
CT phantoms: early history

- Zerhouni et al. developed a phantom for CT density measurement
  - Used as standard for identifying calcified nodules
- Quantitative metric
  - Lesion calcification status

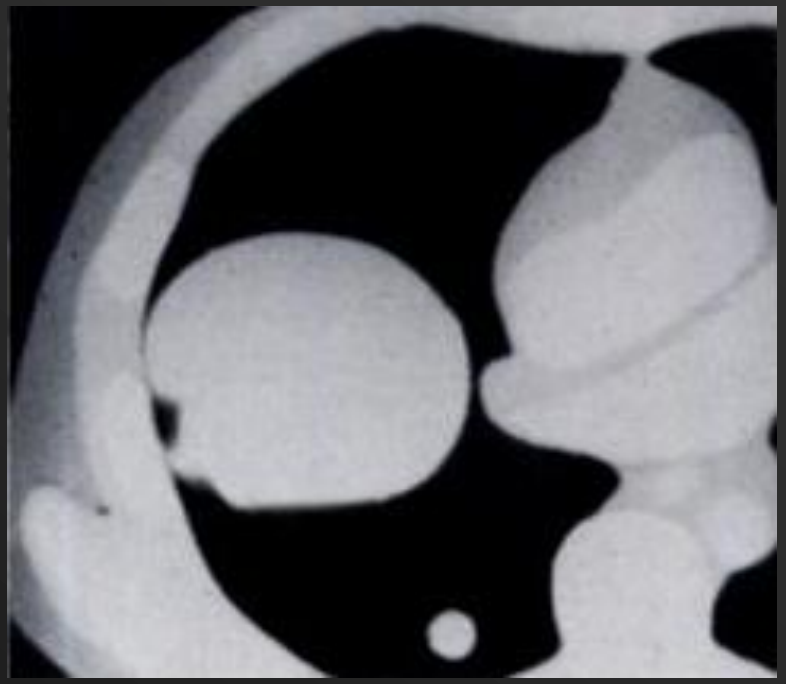


*Zerhouni et al. Radiology. 1983.

CT phantoms: early history



Patient scan of lung nodule

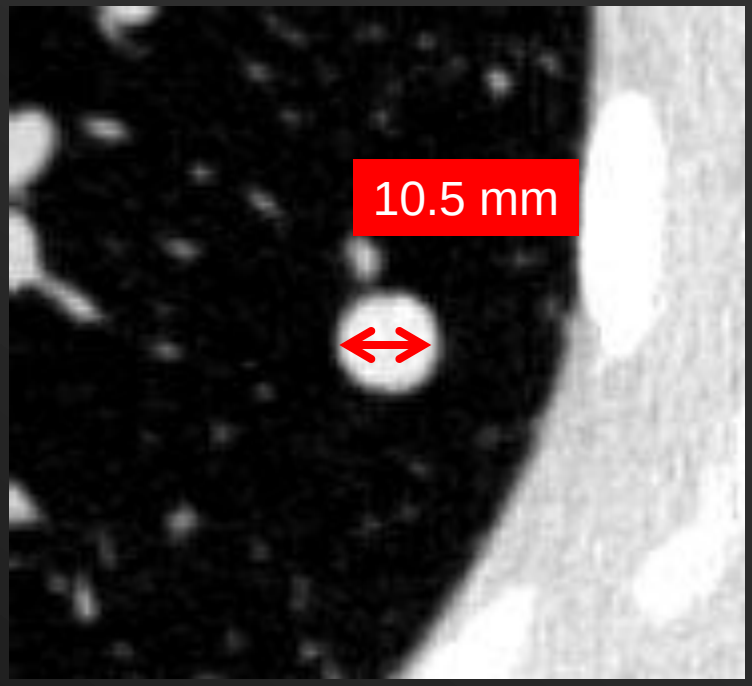


Phantom scan (liver simulated with water sack)

*Zerhouni et al. Radiology. 1983.

CT lesion size as imaging biomarker

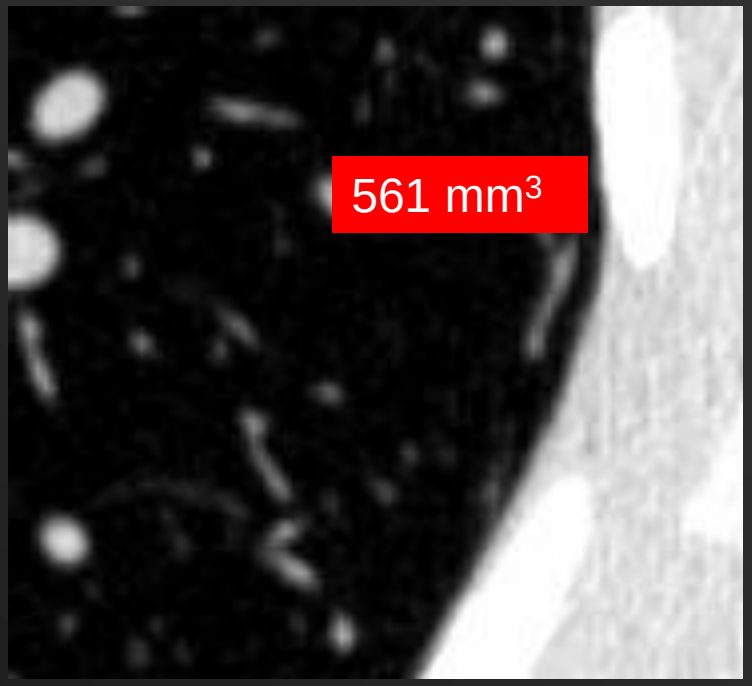
- RECIST
  - Metric
    - 1D, longest in-plane diameter for a lesion
  - Response categories are defined based on measurement uncertainty
    - Complete response
    - Partial response
    - Stable disease
    - Progressive disease
  - Implied symmetric growth assumptions



FDA-018-6591

CT lesion size as imaging biomarker

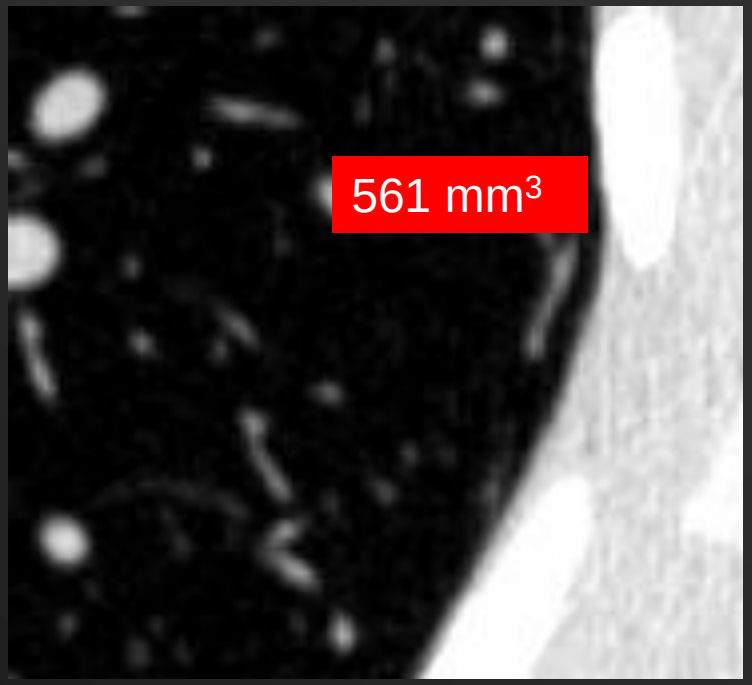
- Lesion volume
  - Metric
    - 3D estimate of lesion size
  - Clinical tools available
    - Generally semi-automated
  - Less common than RECIST in drug trials & clinical practice



FDA-018-6591

CT lesion size as imaging biomarker

- Lesion volume
  - Measurements can be time consuming
  - More difficult estimation task
  - Doesn’t have widely accepted guidelines for use



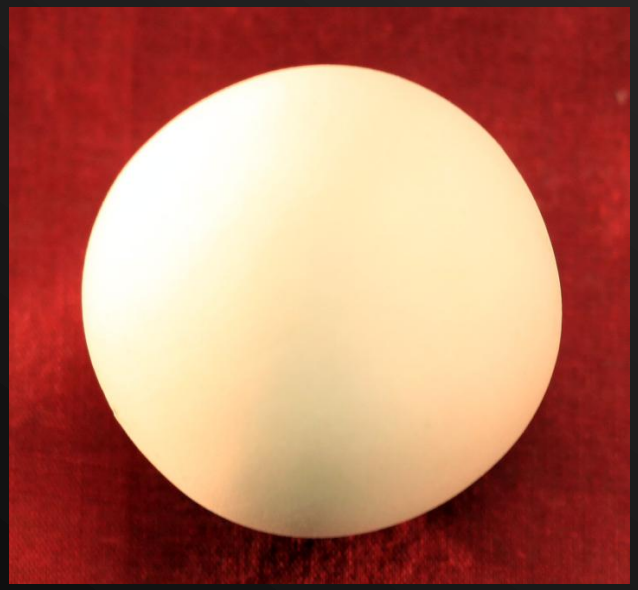
FDA-018-6591

Part 1

Synthetic lesions and phantom designs

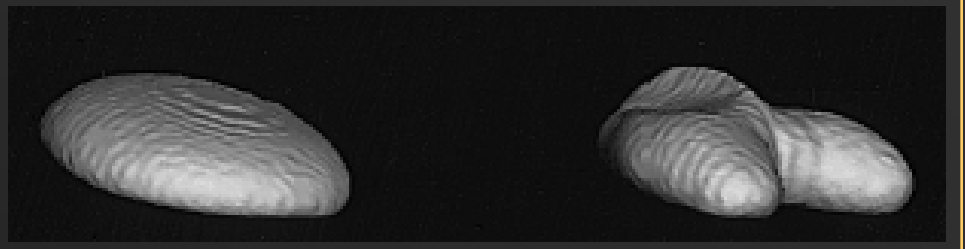
Examples of lesions

- Fixed shape lesions

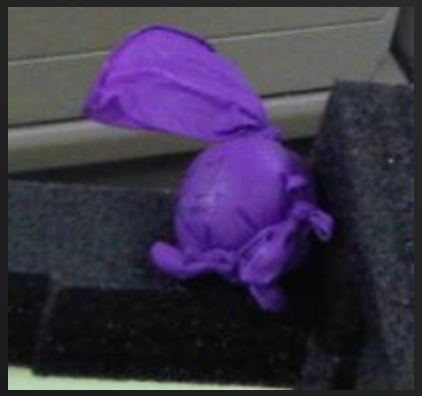


Examples of lesions

- Deformable lesions
  - Same volume, variable shape
  - Materials
    - Deformable silicon
    - Water filled gloves



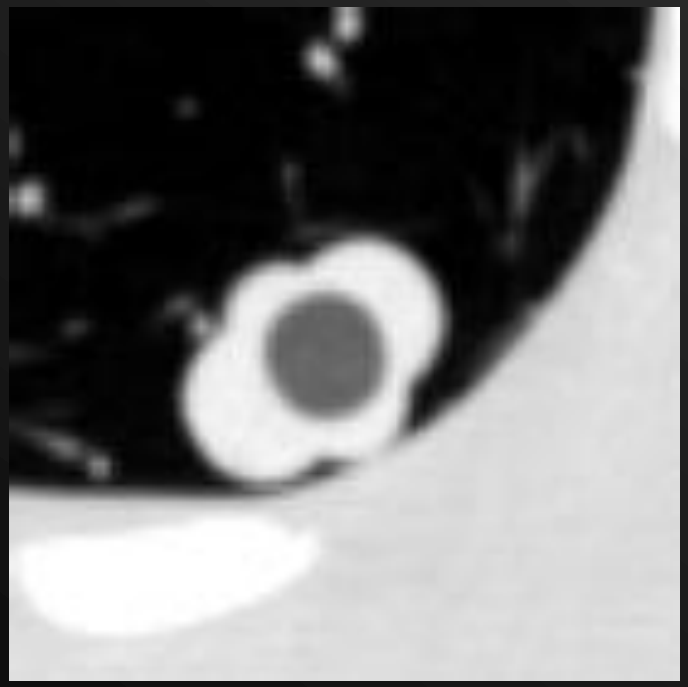
*Yankelevitz et al. Radiology. 2000.



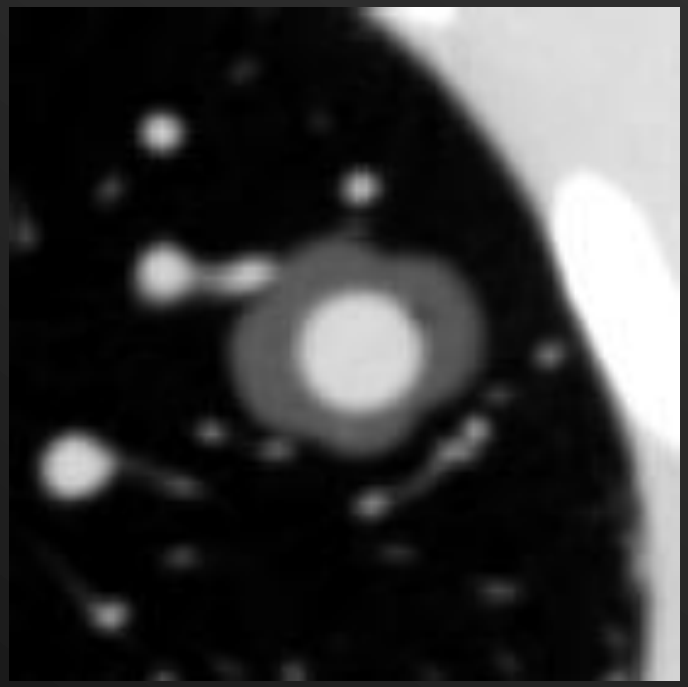
**Huang et al. SSG18-06. RSNA, 2012.

Examples of lesions

- Mixed density lesions



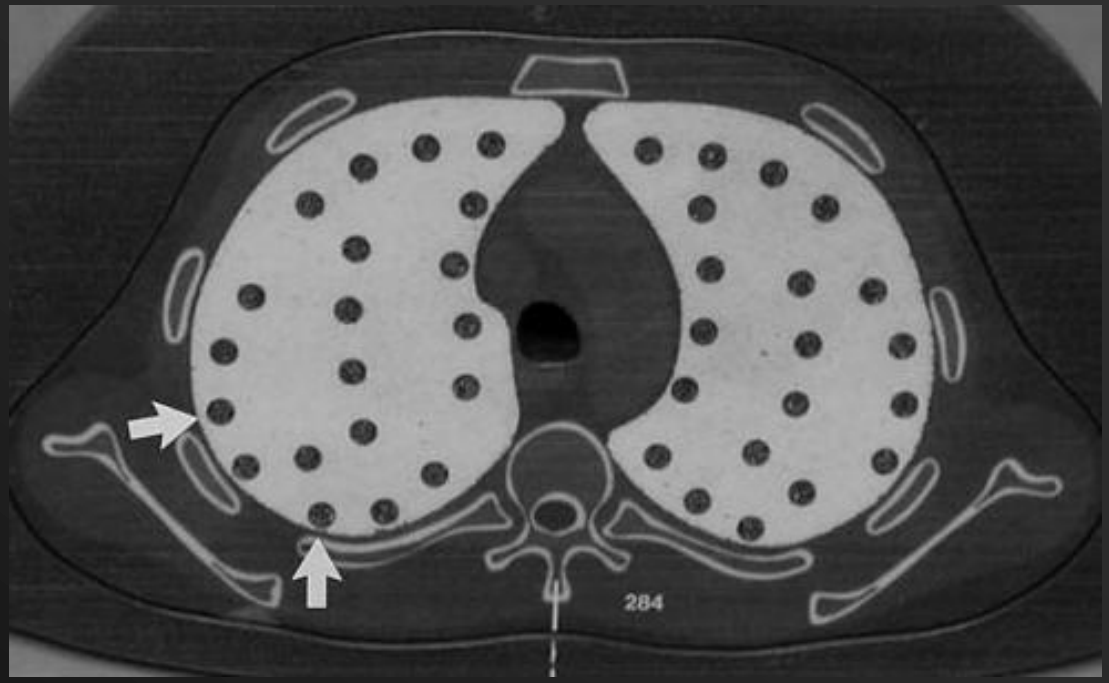
FDA-085-9622



FDA-085-9622

Examples of phantoms

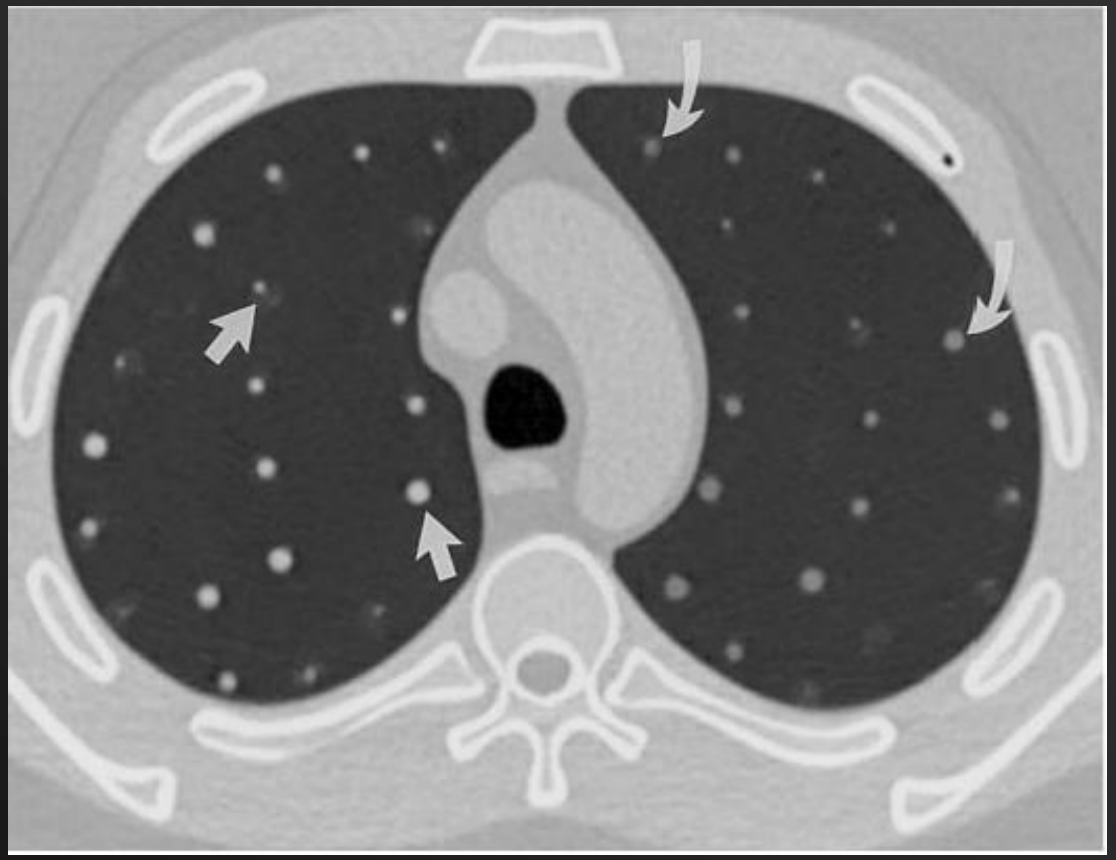
- 8-mm wells
  - 1 nodule inserted within each well surrounded by simulating lung parenchyma
  - Realistic lung background with fine texture



*Ko et al. Radiology. 2003;228:864-70.

Examples of phantoms

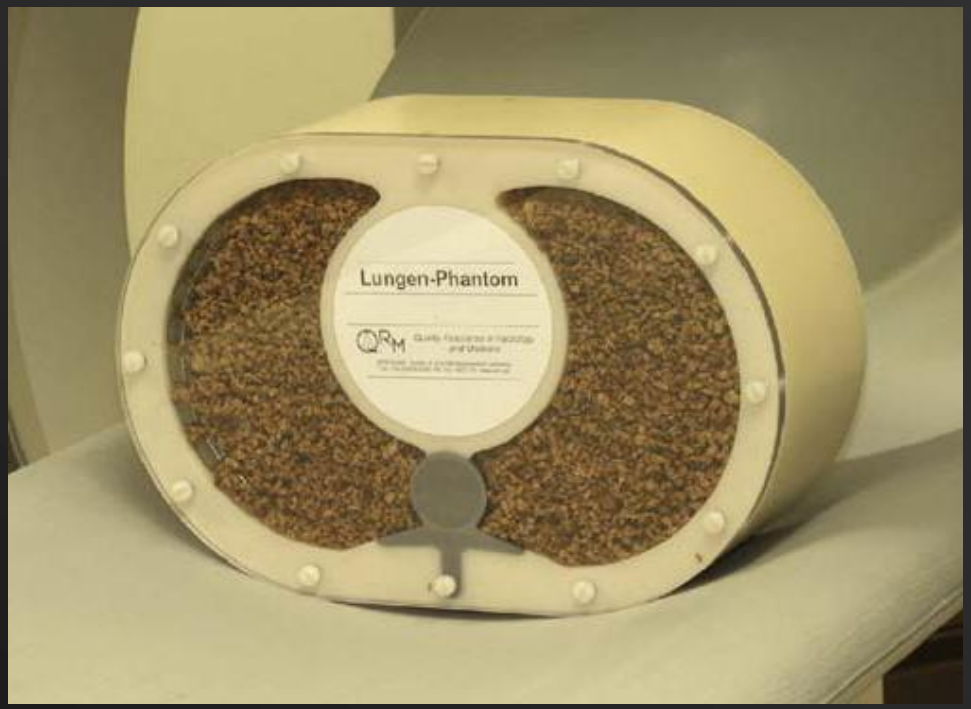
- 8-mm wells
  - 1 nodule inserted within each well surrounded by simulating lung parenchyma
  - Realistic lung background with fine texture



*Ko et al. Radiology. 2003;228:864-70.

Examples of phantoms

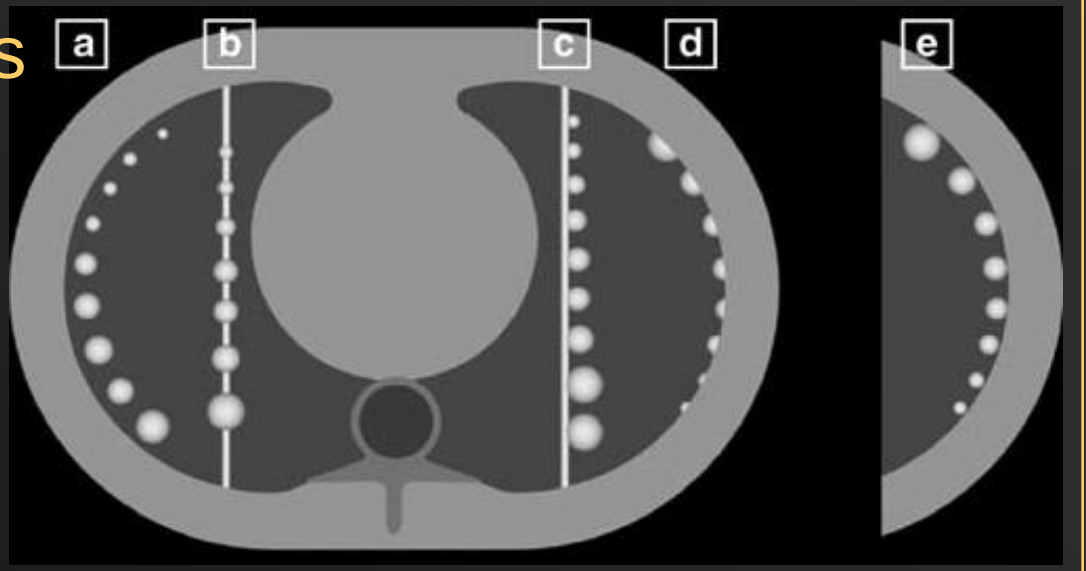
- 5 different types of nodule attachments



*Das et al. Eur-Radiol. 2007;17:1979-84.

Examples of phantoms

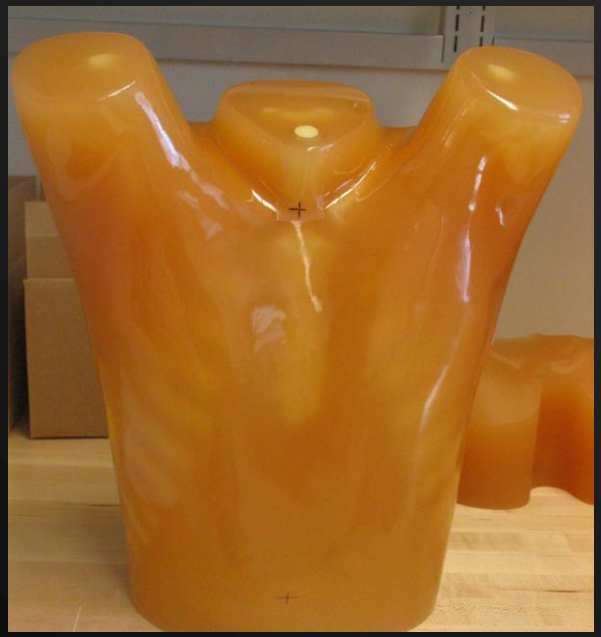
- 5 different types of nodule attachments
  - Isolated nodules
  - Nodules around vessels
  - Vessel attached nodules
  - Pleural nodules
  - Pleural attached nodules

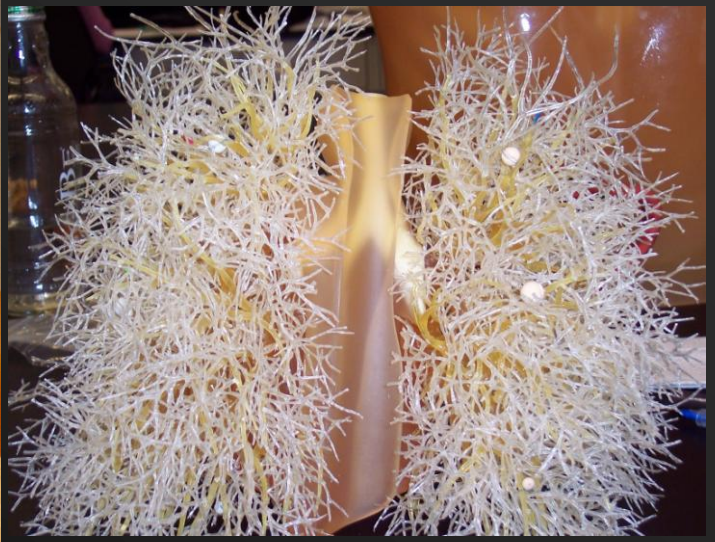


*Das et al. Eur-Radiol. 2007;17:1979-84.

Examples of phantoms

- Anthropomorphic phantoms with vascular structure

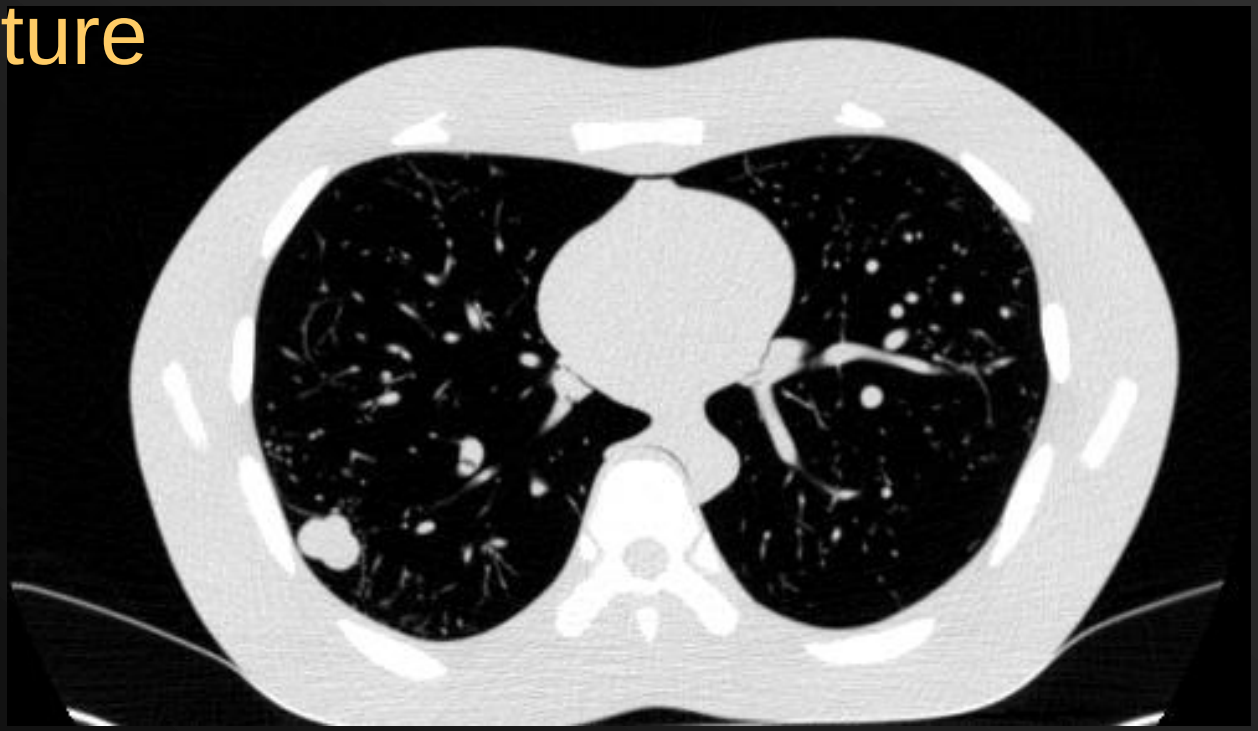




*Kyotokagaku Incorporated, Tokyo, Japan

Examples of phantoms

- Anthropomorphic phantoms with vascular structure



*QIBA vCT Project. 2011.

Part 2

Performance metrics

Common performance metrics

- Size estimates
  - Simple error

$$\text{Error} = Sz_m - Sz_{True}$$

  - Percent error (Petrick et al, SPIE. 2011)

$$PE = 100x\frac{Sz_m - Sz_{True}}{Sz_{True}}$$

  - Absolute size error (Ko et al. Radiol. 2003)

$$AE = \left|Sz_m - Sz_{True}\right|$$

  - Absolute percent size error (Das et al. ERadiol. 2007)

$$APE = 100x\frac{\left|Sz_m - Sz_{True}\right|}{Sz_{True}}$$

Common performance metrics

- Accuracy
  - Mean error

    $\text{Bias} = \text{mean}(\text{Error}) = E(\text{Error})$
  - Mean percent error

    $\text{Percent Bias} = \text{mean}(PE) = E(PE)$

    Allows for true bias estimation
  - Mean absolute error

    $\text{Mean}(AE) = E(AE)$
  - Mean absolute percent size error

    $\text{Mean}(APE) = E(APE)$

    Over/under estimates weighted equally

Common performance metrics

- Precision
  - Error

$$\text{Std}(\text{Error}) = \text{Std}(Sz_M - Sz_{True})$$

  - PE

$$\text{Std}(PE) = \text{Std}\left(100x\frac{Sz_m - Sz_{True}}{Sz_{True}}\right)$$

  - AE

$$\text{Std}(AE) = \text{Std}(|Sz_M - Sz_{True}|)$$

  - APE

$$\text{Std}(APE) = \text{Std}\left(100x\frac{|Sz_m - Sz_{True}|}{Sz_{True}}\right)$$

Common performance metrics

- Change in size between scans
  - Simple change

$$\text{Change} = Sz_2 - Sz_1$$

  - Percent change

$$PC_1 = 100x\frac{Sz_1 - Sz_2}{Sz_1}, PC_m = 100x\frac{Sz_1 - Sz_2}{\overline{Sz}}$$

  - Absolute change

$$AC = |Sz_2 - Sz_1|$$

  - Absolute percent change

$$APC_1 = 100x\frac{|Sz_1 - Sz_2|}{Sz_1}, APC_m = 100x\frac{|Sz_1 - Sz_2|}{\overline{Sz}}$$

Common performance metrics

- Change in size between scans
  - Doubling time
    - Has clinical meaning but not tractable statistically
    - Zero change $\rightarrow \infty$ doubling time
    - These types of metric should be avoided in evaluation

Scaling data

- When/how should data be scaled to improve comparisons?
  - How to compare 1D with 3D volume sizing?
    - Data normalization

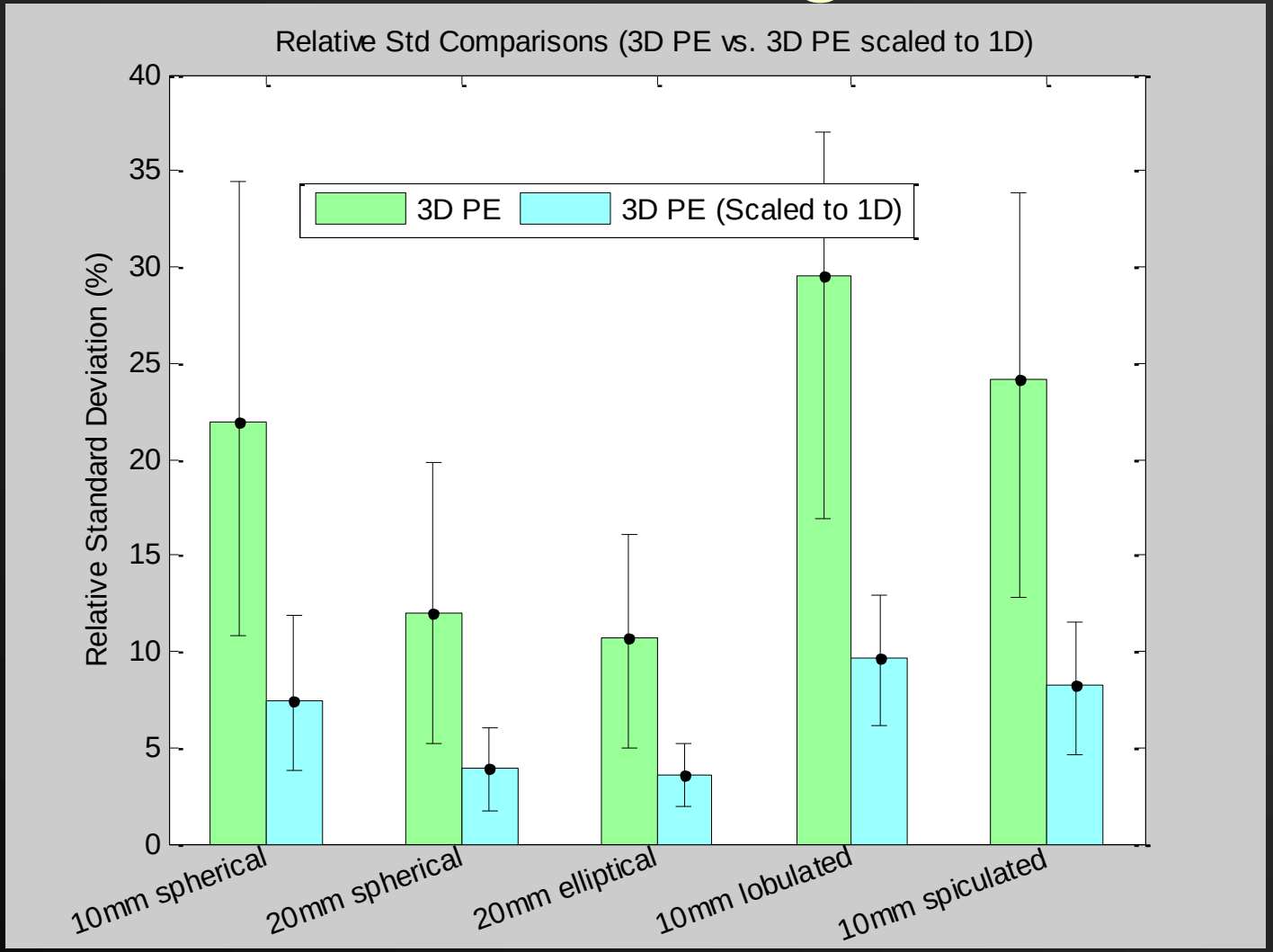
$$PE = \frac{Sz_m - Sz_{True}}{Sz_{True}} \times 100\%$$

    - Data scaling, normalization

$$PE^1_{Size} = \frac{Size^1_{Est} - Size^1_{True}}{Size^1_{True}} \times 100\%, \quad PE^3_{Size} = \frac{\sqrt[3]{Size^3_{Est}} - \sqrt[3]{Size^3_{True}}}{\sqrt[3]{Size^3_{True}}} \times 100\%$$

    - Statistical normalization
      - Log transformation

Different scaling choice

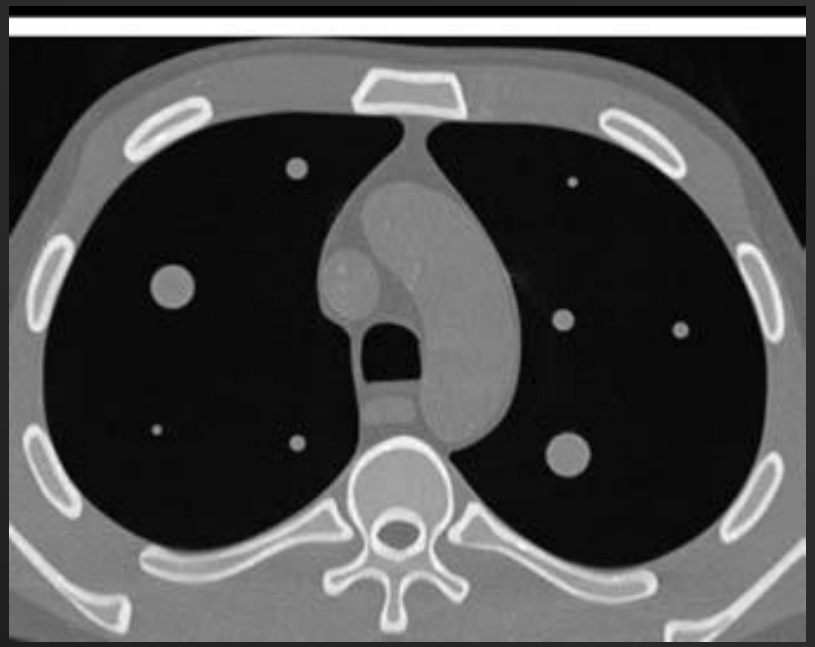


Part 3

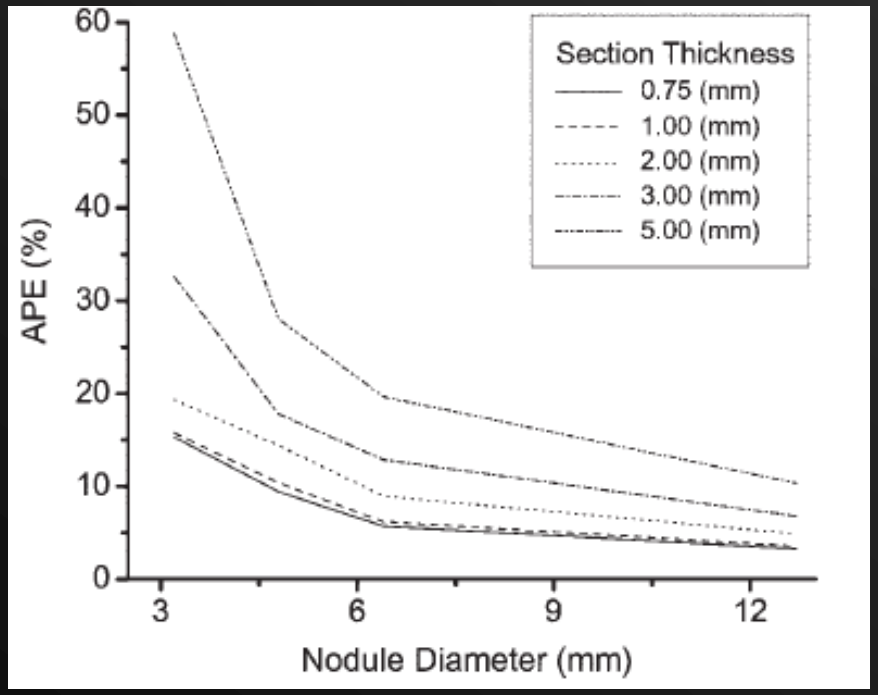
Lessons from the literature, QIBA, FDA studies

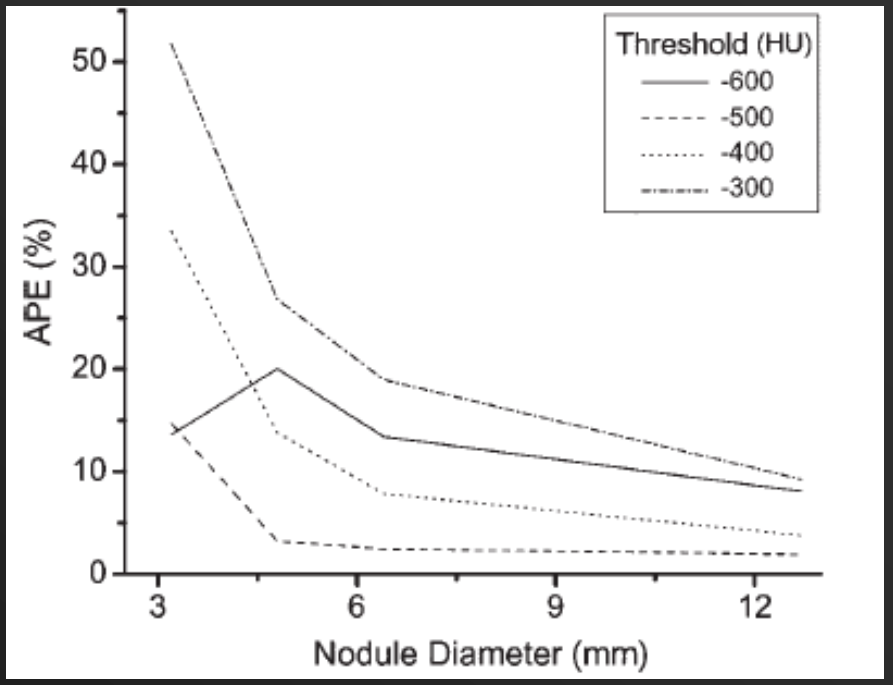
Goo et al. Radiology. 2005

- CT scans of nodules
  - Nodules (130 HU)
    - 3.2, 4.8, 6.4 & 12.7 mm diameter spheres
  - Scanner
    - Siemens Somatom Sensation 16
    - Reconstructed slice thicknesses
      - 0.75, 1.0, 2.0 3.0 and 5.0 mm
    - 3 FOV's
      - 0.20, 0.39, 0.59 mm in-plane resolution
  - Volume measurements
    - Rapidia commercial tool
    - Thresholds: -300, -400, -500, -600 HU



Goo et al. Radiology. 2005

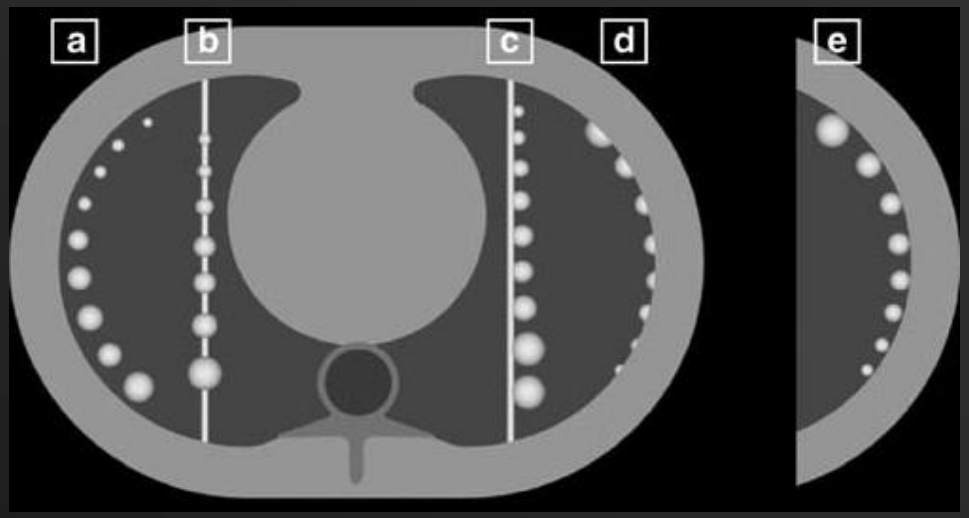




- Measurement error strong function of
  - Slice thickness ($p<0.01$)
  - Segmentation threshold ($p<0.02$)
- FOV not found to have a statistically significant impact ($p>0.05$)

Das et al. Eur Radiol. 2007

- Nodules (35 HU)
  - 3.0 – 10 mm diameter spheres
  - 5 different attachment types
- Scanners
  - Siemens Somatom Sensation 16
  - GE LightSpeed Pro 16
  - Philips Briliance 16
  - Toshiba Aquilion 16
  - ~0.75 & ~1.5 mm slice thicknesses
  - Low-dose and standard-dose exposures
- Volume measurements
  - Prototype semi-automated Siemens LungCARE tool
    - Users define seed marker



Das et al. Eur Radiol. 2007

Mean(APE) ± SD		Siemens 16 1.5 LD	GE 16 1.25 LD	Philips 16 1.5 LD	Toshiba 16 1.0 LD
Nodule Types	Isolated	9.8±8.0	18.1±8.1	15.8±9.5	11.2±7.3
	Through vessel	5.0±1.9	10.2±14.6	5.8±3.5	4.8±6.7
	Attached to Vessel	4.5±3.1	9.8±10.4	2.6±2.6	3.3±2.1
	Pleural	19.0±7.8	10.0±6.9	15.0±9.4	8.2±6.2
	Pleural Attached	9.4±2.7	17.4±11.3	8.8±3.3	6.2±1.7

- Nodule location/connection influence measurements

Das et al. Eur Radiol. 2007

Mean(APE) ± SD		Siemens 16 1.5 LD	GE 16 1.25 LD	Philips 16 1.5 LD	Toshiba 16 1.0 LD
Nodule Types	Isolated	9.8±8.0	18.1±8.1	15.8±9.5	11.2±7.3
	Through vessel	5.0±1.9	10.2±14.6	5.8±3.5	4.8±6.7
	Attached to Vessel	4.5±3.1	9.8±10.4	2.6±2.6	3.3±2.1
	Pleural	19.0±7.8	10.0±6.9	15.0±9.4	8.2±6.2
	Pleural Attached	9.4±2.7	17.4±11.3	8.8±3.3	6.2±1.7

- CT vendor influence measurements

Das et al. Eur Radiol. 2007

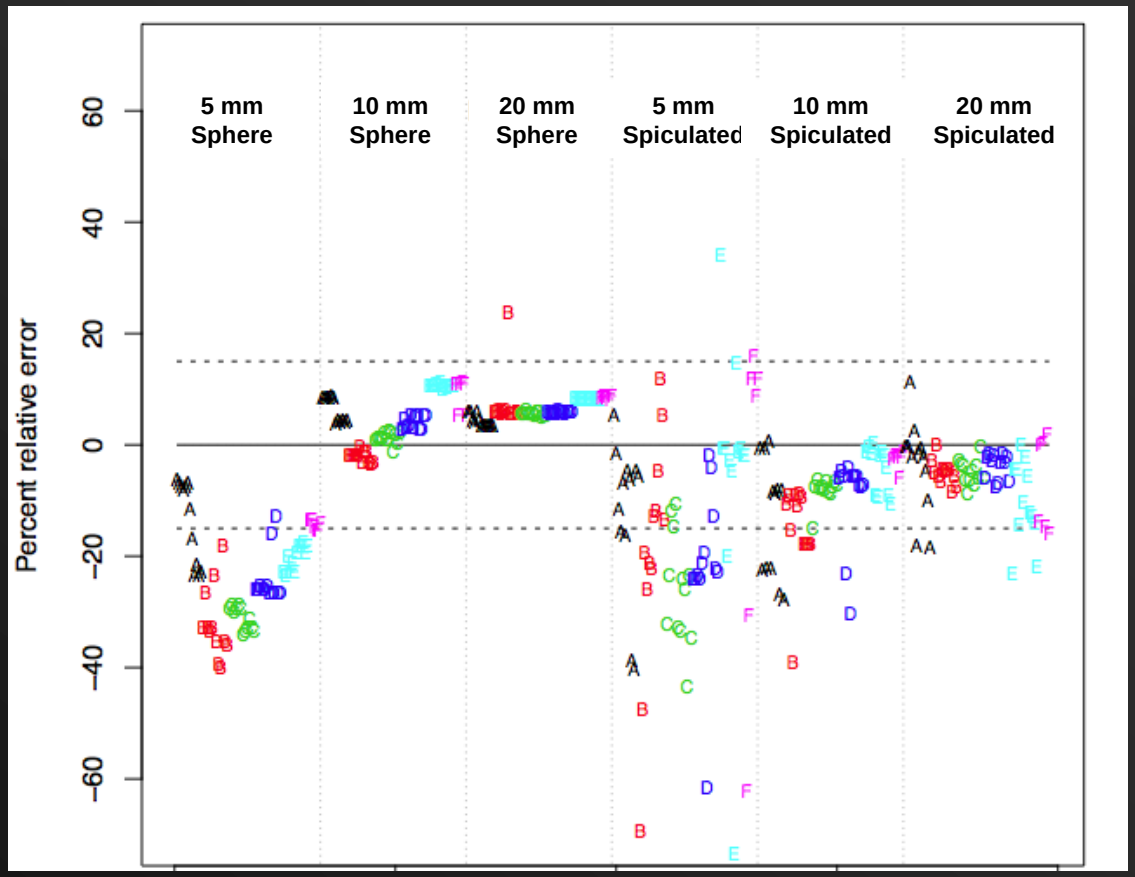
- Factor influencing volume measurements
  - Nodule location/connection type
  - CT vendor (p=0.004)
  - CT parameters
    - Collimation (p=0.021)
    - Slice thickness (p=0.019)
    - Dose (p=0.099, not significant)

QIBA vCT 1C subproject

- Objective
  - Evaluate if 3D measurements are reproducible across scanners
- Six scanners at six different sites
  - Phillips 16 (2 sites), Philips 64, Siemens, GE 64, Toshiba 64
- Six nodules
  - Spherical (5, 10, 20 mm)
  - Spiculated (5, 10, 20 mm)

QIBA 1C Results

- A-F are the different scanners
- Scanner can have impact
  - Nodule type is important interaction

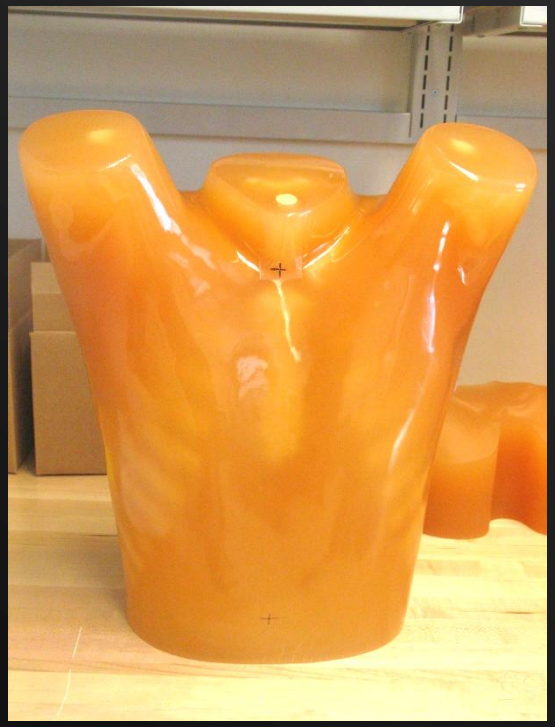


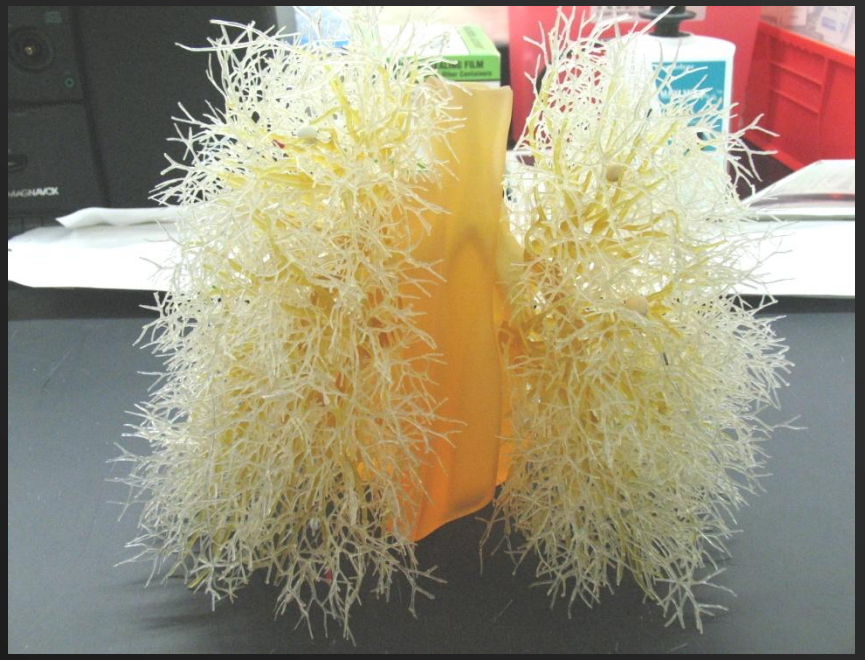
*Courtesy of Chuck Fenimore, QIBA 1C leader

QIBA vCT 1A subproject

- Objective
  - To estimate bias/variance of radiologists estimating the size of nodules from CT scans of an anthropomorphic phantom

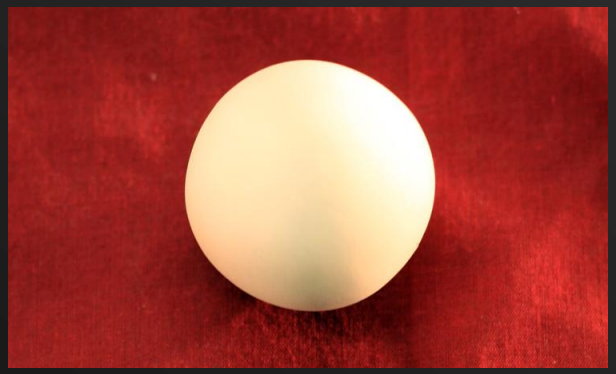
Dataset (Thorax Phantom)



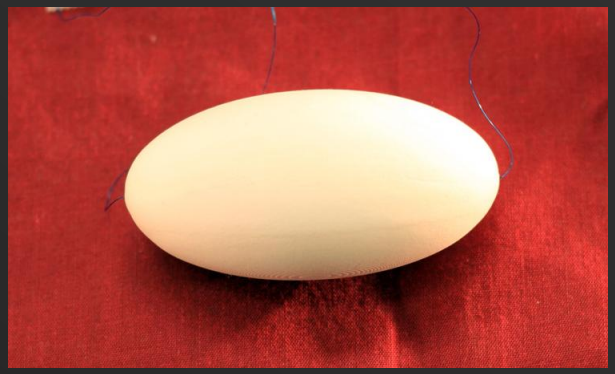


Anthropomorphic thorax phantom (Kyotokagaku Incorporated, Tokyo, Japan)

Nodules

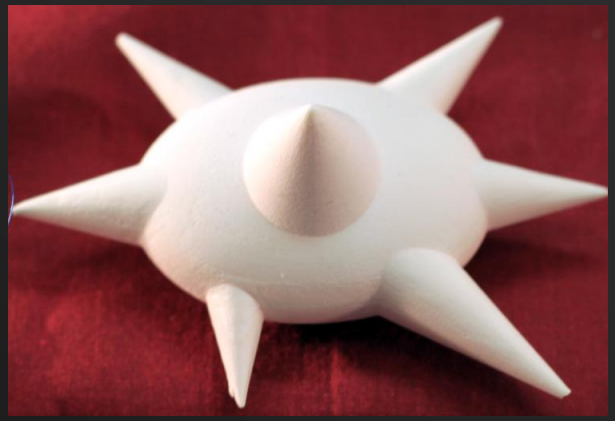


Spherical



Elliptical

Lobulated



Spiculated

Dataset (Nodules)

- 10 nodules attached to lung vasculature

Nodule Shape	Equivalent Diameter	CT Densities
Spherical	10 mm	-10 HU, +100 HU
Spherical	20 mm	-10 HU, +100 HU
Elliptical	20 mm	-10 HU, +100 HU
Lobulated	10 mm	-10 HU, +100 HU
Spiculated	10 mm	-10 HU, +100 HU

Dataset (Scanning)

Acquisition Parameter	Values
Scanner	Philips 16-slice Mx8000 IDT scanner
Tube Voltage	120 kVp
Exposure	100 mAs/slice
Pitch	1.2
Reconstructed Slice Thickness	0.8 mm (0.4 mm interval,16x0.75 mm coll) 5.0 mm (2.5 mm interval, 16x1.5 mm coll)
Reconstruction Kernel	Detail
Repeat Exposures	2 repeat scans of each nodule

40 dataset evaluated: 10 nodules X 2 slice thickness X 2 repeat scans

Reading Protocol

- 6 readers measured size of 40 nodules using 3 size measurement techniques in each of two reading sessions
- Measures:
  - Manual uni-dimensional largest in-slice diameter measure (1D)
  - Manual bi-dimensional largest in-slice area measure (2D)
  - Semi-automated 3D volume software (3D)
- Analysis
  - Bias and variance of size estimates
  - Reference standard
    - Longest physical dimension of nodule
    - Largest cross-sectional area of nodule
    - Volume

Data Normalization

- All size data was scaled to 1D and normalized to facilitate comparison

$$PE^{1}_{Size} = \frac{Size^{1}_{Est} - Size^{1}_{True}}{Size^{1}_{True}} \times 100\%,$$

$$PE^{2}_{Size} = \frac{\sqrt{Size^{2}_{Est}} - \sqrt{Size^{2}_{True}}}{\sqrt{Size^{2}_{True}}} \times 100\%,$$

$$PE^{3}_{Size} = \frac{\sqrt[3]{Size^{3}_{Est}} - \sqrt[3]{Size^{3}_{True}}}{\sqrt[3]{Size^{3}_{True}}} \times 100\%$$

Performance Metrics

- Relative bias

$$Bias^{i}_{Rel}\left(PE^{i}_{Size}\right) = \mathrm{E}[PE^{i}_{Size}] - 0$$

- Relative standard deviation

$$Std^{i}_{Rel}\left(PE^{i}_{Size}\right) = std\left(PE^{i}_{Size}\right)$$

ANOVA & Goodness of Fit

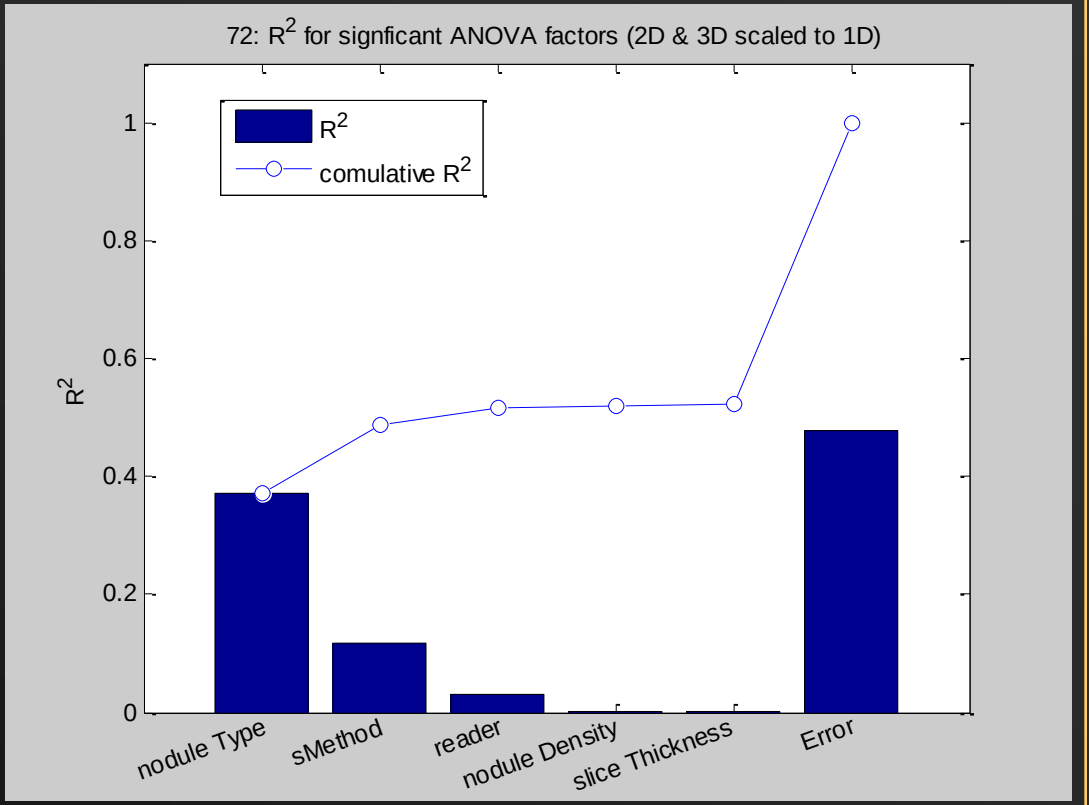
- Used to identify the most important contributing factors and interactions to include in our bias analysis
  - Goodness of fit defined by R^2

1-way ANOVA & Goodness of Fit

- Factors
  - Nodule type*
  - Sizing method*
  - Reader*
  - Nodule density*
  - Slice thickness*

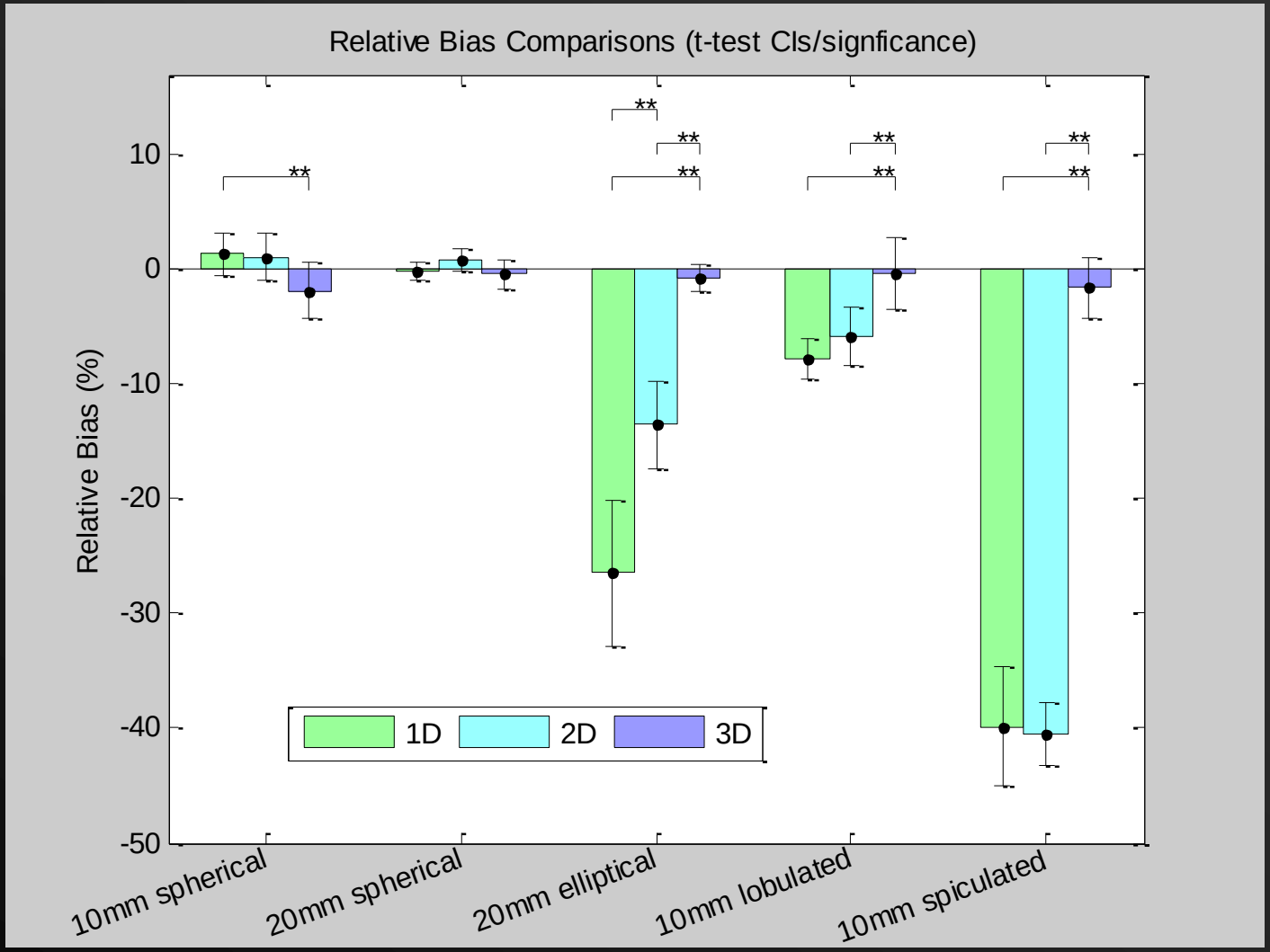
  - Nodule Set (p=0.7)
  - Reading Sess (p=0.7)

- Reader, Nodule Density, Slice Thickness explained little variation

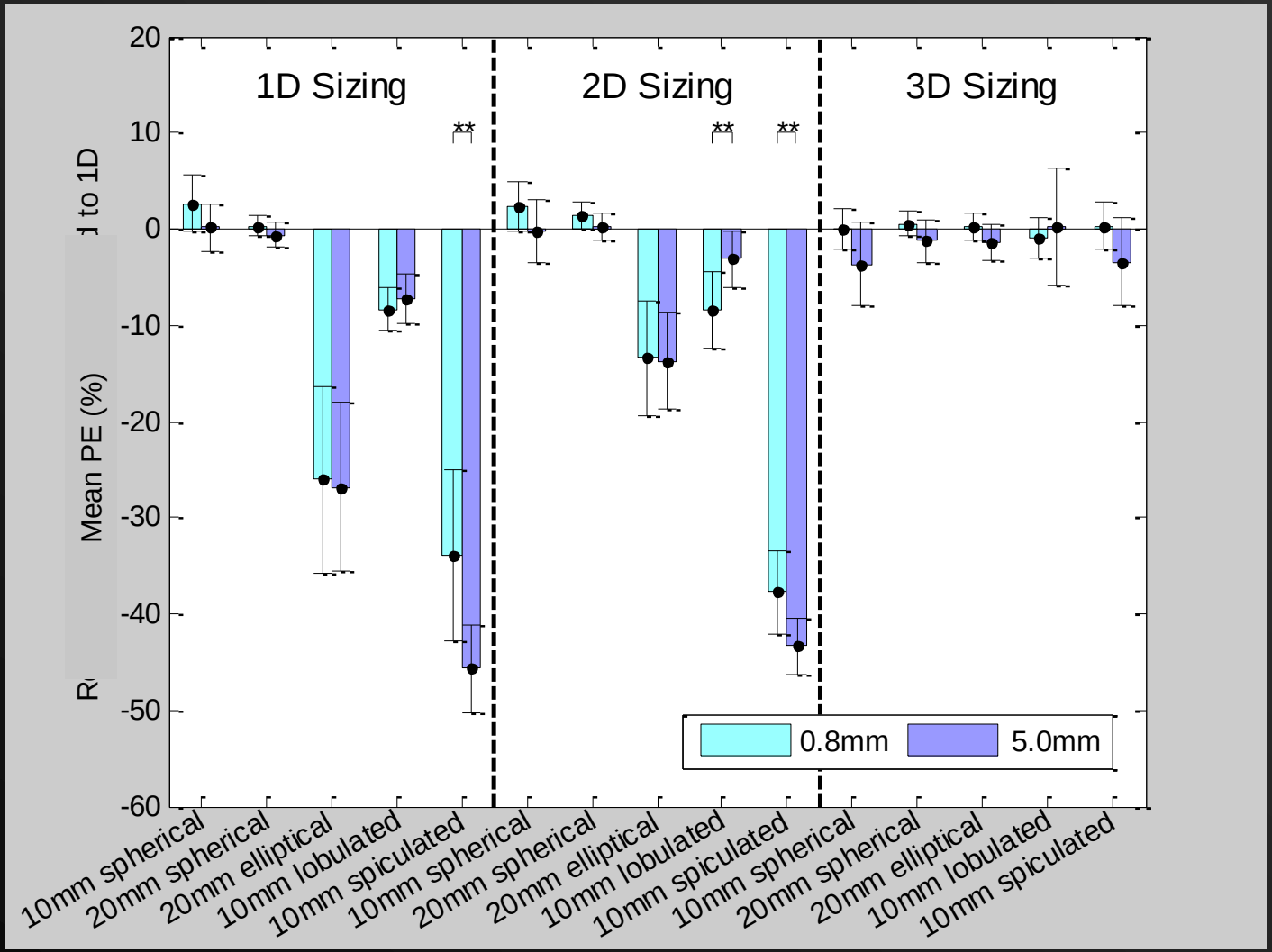


* Significant effects

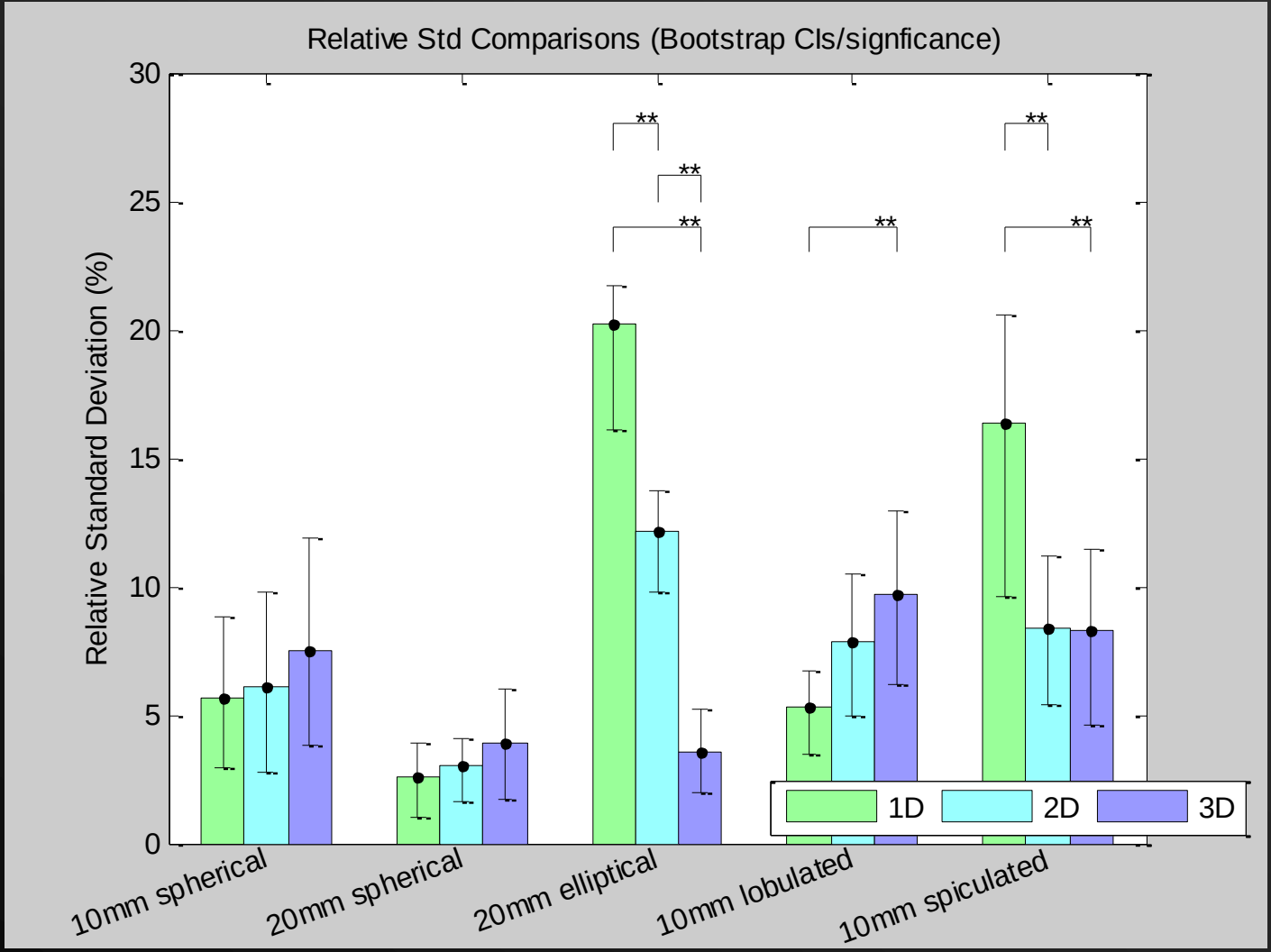
Comparison of mean PE (bias)



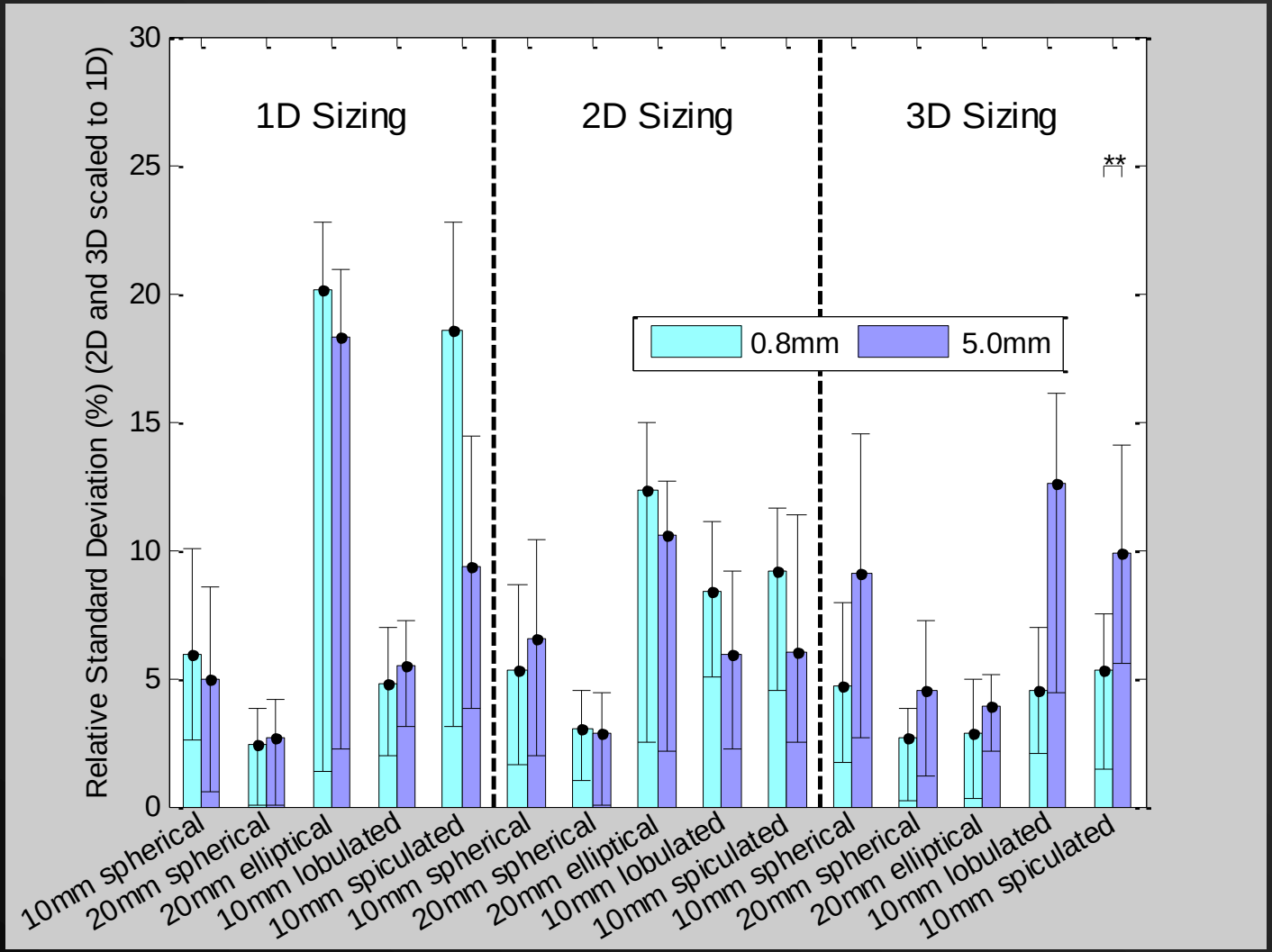
Mean PE (bias) by slice thickness



Comparison of Variances (Std PE)



Variance (Std PE), func slice thickness



Summary

- CT dose doesn’t appear to substantially affect bias in size estimation
  - Lower dose does increase variability
- Nodule characteristics & measurement technique play a role but this is task dependent
  - QIBA 1A: Slice thickness played a large effect for volume sizing but less effect for 1D and 2D sizing

Summary

- Phantom studies can help in understanding interactions among multiple factors including
  - CT acquisition factors
  - CT reconstruction factors
  - Measurement tool factors
  - Nodule/lesion factors
- Phantom results may serve as bound on achievable clinical performance

Ongoing phantom-based research

- Developing consensus performance metrics and analysis methodologies for size estimation evaluation
  - QIBA Metrology group drafting a set of manuscript to help develop consensus approaches for assessing technical performance for QIBs.
- Ongoing studies investigating
  - Inter-comparison of volume estimation algorithms
  - Nodule growth over time
  - …