2015 Report from the CT Volumetry Biomarker Committee of the Quantitative Imaging Biomarkers Alliance

Gregory Goldmacher, MD, PhD, MBA; Samuel Armato III, PhD; Maria Athelogou, PhD; Andrew Buckler, MS; James Mulshine, MD; Nicholas Petrick, PhD; Berkman Sahiner, PhD; Ehsan Samei, PhD; Jenifer Siegelman, MD, MPH

Quantitative Imaging Biomarkers Alliance RSNA

Background And Previous Work

Goal: Define Analytic Performance of 3D CT Volumetry as a Biomarker for Tumor Response

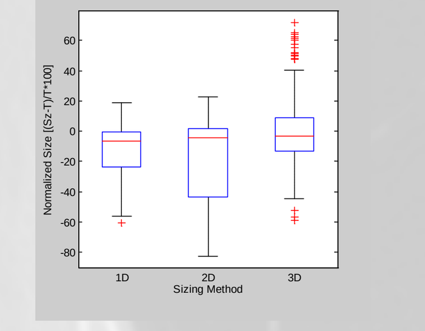
Stakeholders: FDA, NIST, NCI, FNIH, ACRIN, manufacturers and developers, CROs, PINTAD, and academic centers.

Sizing of Synthetic Spherical and Non-Spherical Lung Nodules (aka "1A")

Aim

To estimate bias/variance of radiologists estimating the size of synthetic nodules

Results



- Box and whisker plot of reader size estimates for the 1D, 2D and 3D sizing methods (all phantom nodules included).

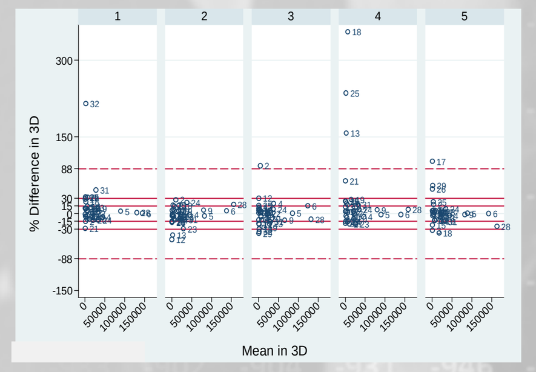
Summary

- Overall, 3D method provided low bias estimates of nodule volumes
- 3D method applied to thin slice data provided low bias & low variance estimates

Minimum Detectable Change in Clinical Trial Workflows (aka "1B")

Hypothesis The minimal detectable change in tumor size will be smaller when using a side by side ("clinical trial workflow") review setting than when using timepoints presented in random order.

Results



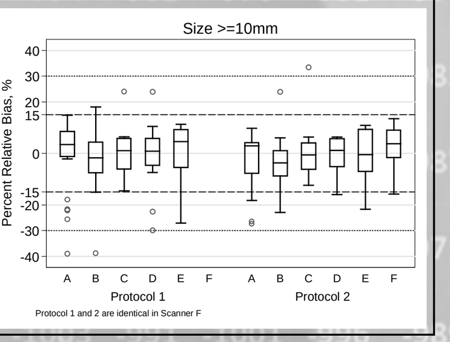
Discussion

1. Measurement variability is considerably reduced when using the locked, sequential read approach compared to randomized timepoint reads
2. Should inform the QIBA profile as to "best practices" for clinical trials

Evaluating inter CT scanner effects in clinician sizing of phantom nodules (aka "1C")

Aim: characterize accuracy and precision in reader measurements of volumes of six phantom nodules collected on six scanners.

Results



Conclusion

- Relative bias is within a tolerance of 15%.
- Scanner equivalence is found only for the larger synthetic lesions (10 mm and 20 mm). This finding confirms the lesion sizing guidance (10 mm and up) in QIBA CT imaging profile.
- Equivalence of the protocols supports the imaging protocol as used by ACRIN Trial 6678.

Advanced Disease Profile

Profile structure:

Section 1: Executive Summary
Section 2: Claim – What performance can be achieved if profile is followed
Section 3: Activities – Biomarker activity specifications to achieve the performance claim
Section 4: Assessment Procedures – Procedures for assessing compliance with specifications

Current draft on the QIBA wiki: http://qibawiki.rsna.org/index.php?title=CT_Volumetry_Biomarker_Ctte

Recent work:

- Completed public review; Revisions to Claim (balancing simplicity, clinical utility, and statistical rigor)
- Revisions to clinical interpretation, assessment procedures for scanner and analysis software, image QA
- Profile implemented in limited settings.

Next step: "FIELD TEST" TECHNICAL VALIDATION

Jenifer Willmann Siegelman MD, MPH
Brigham and Women's Hospital / Harvard Medical School / Partners Health Care

Aims:

- Assess performance of CT Volumetry Profile for solid tumors in vivo
- Compare inter-scanner measurement variability (across up-to-date platforms)
- Collect data that can be sequestered for compliance testing

Method: QIBA CT Volumetry Profile to be executed on human subjects in a multi-vendor, multi-center trial

Study Design: A test-retest in which clinical subjects with known measurable tumors in lungs, liver, and lymph nodes will be scanned two times on the same day without contrast, or on subsequent days with low dose non-ionic contrast. Subjects will be randomized to same scanner, or alternate scanner. Image data will be assessed by human readers using a variety of software algorithms to measure tumor volumes

Results: The results will be represented in a table as shown below.
This work will close critical gaps in uncertainty about the precision of measurement by CT. This knowledge will improve clinical care through more accurate assessment of the "no change" state, and allow smaller changes in tumor size (growth or shrinkage) to be measured with confidence, making them more relevant in clinical care and as research endpoints.

	Different Scanner				Same Scanner			
	Different software		Same software		Different software		Same software	
	Different Reader	Same Reader	Different Reader	Same Reader	Different Reader	Same Reader	Different Reader	Same Reader
Precision								

Various QIBA projects and activities have been funded in whole or in part with Federal funds from the National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Department of Health and Human Services, under Contract Nos. HHSN268201000050C and HHSN268201300071C.

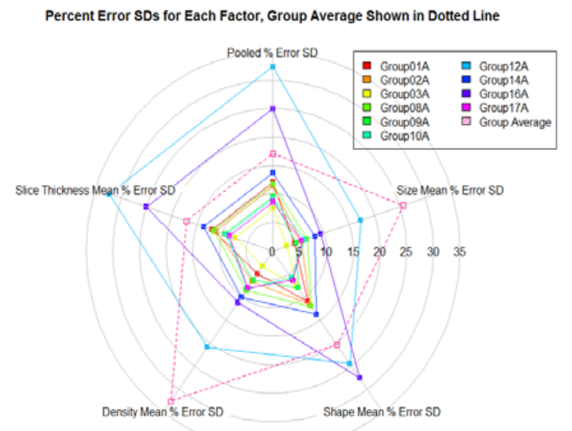
Recent And Ongoing Research On Advanced Disease

Algorithmic Volume Quantification (3A) Group

Inter-algorithm Performance Study Using FDA Phantom Data

Paper undergoing review

Aim: Estimate absolute volumes using CT phantom data. Report bias and variance.



Study Results :

Percent error for all participants

Using only nodules that met QIBA CT profile requirements, the standard deviation from pooled data for all 10 participants are shown by the dotted pink polygon. The pooled standard deviation of all 10 participants is shown as polygons of various colors.

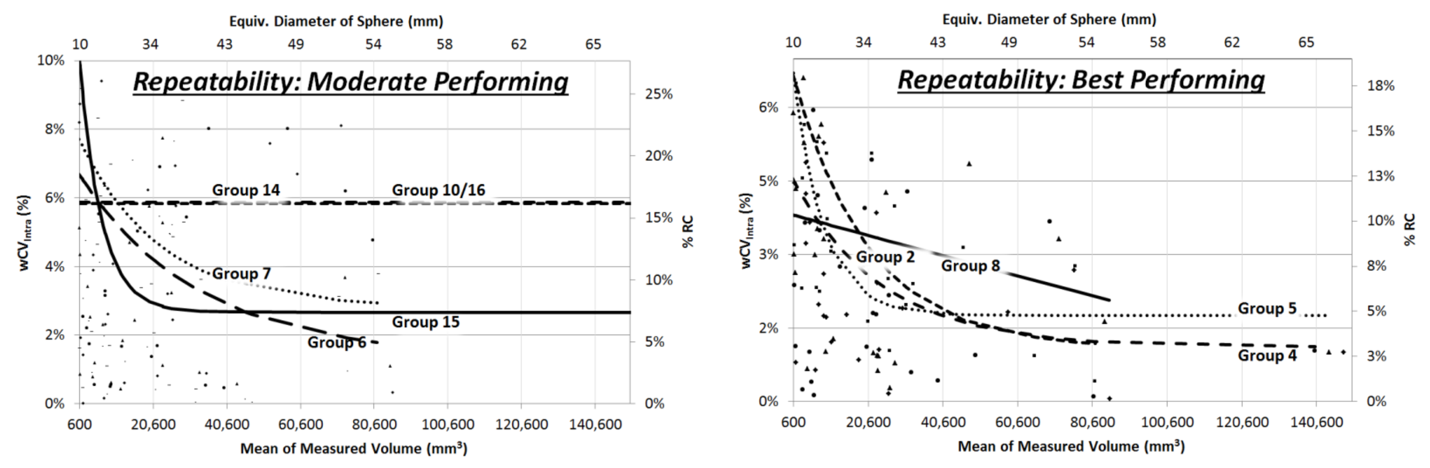
Conclusion: The results support QIBA performance claims that profile-conformant measurements produce results where the 68% confidence interval for the systematic deviation between an estimate and the true value (i.e., +/- 1-sigma) is less than 15%. Results also address the hypothesis that performance claims for tumor volume may be met by various measurement algorithms ranging from semi- to fully automated methods.

Inter-method Performance Study of Tumor Volumetry Assessment on Computed Tomography Test-retest Data

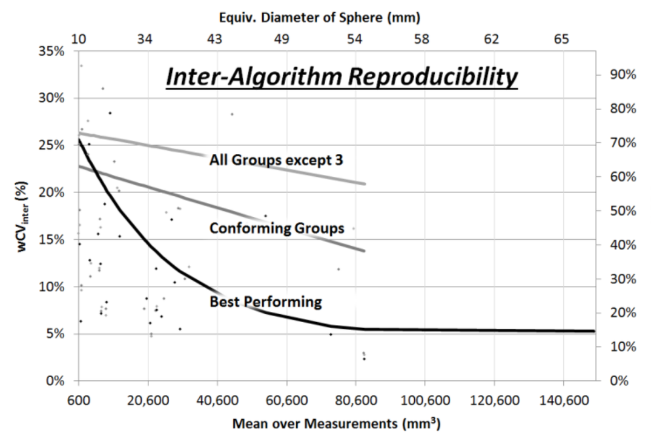
Paper accepted: Journal of Academic Radiology

A. J. Buckler, J. Danagoulian , K. Johnson, S St. Pierre, A. Peskin , M. A. Gavrielides , N. Petrick, N. A. Obuchowski , H. Beaumont, L. Hadjiiski, R. Jarecha, J. M. Kuhnigk , N. Mantri, M. McNitt-Gray, J. H. Moltz , G. Nyiri , S. Peterson , P. Tervé , C. Tietjen, E. von Lavante, X. Ma , M. Athelogou

Industry and academic groups participated in a challenge study. 31 lung cancer test-retest cases were analyzed by 12 participants. Intra-algorithm repeatability and inter-algorithm reproducibility were estimated. Relative magnitudes of various sources of variability were estimated using a linear mixed effects model. Segmentation boundaries were compared to provide a basis on which to optimize algorithm performance for developers.



Intra-algorithm repeatability as a function of measured tumor size. The line fits following exponential functions. Fits for the least performing algorithms could not be made given highly variable results. Left panel shows fit lines for moderate performing algorithms, and right panel for best performing algorithms. The fit lines are truncated where they would imply better performance than the sparse set of points at high tumor volumes actually suggest. RC, repeatability coefficient; wCV_{intra}, within-tumor coefficient of variability.



Inter-algorithm reproducibility analysis across tumor size range. Line fits follow exponential functions. Fit lines are truncated where they would imply better performance than the sparse set of points at high tumor volumes actually suggest. RDC, reproducibility coefficient; wCV_{intra}, within-tumor coefficient of variability.

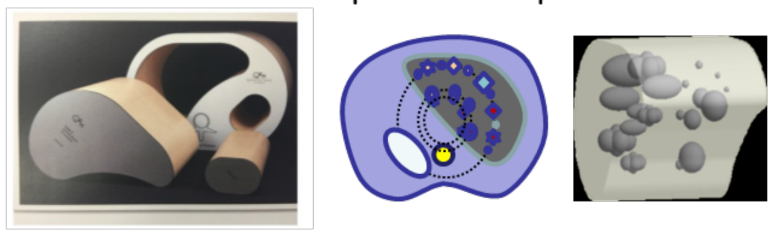
Conclusions:

Nine of the twelve participating algorithms performed at a level sufficient for QIBA conformance on the basis of intra-algorithm repeatability as judged on this data set. Based on these results, change in tumor volume can be measured with confidence to within ±14% using any of the nine conformant algorithms down to tumor sizes of 10 mm or greater.

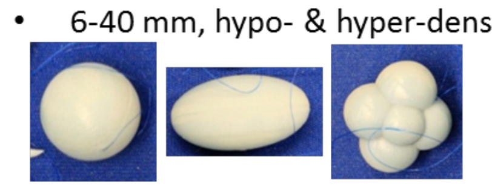
No partition of the algorithms demonstrated sufficiently low reproducibility to meet QIBA requirements for interchangeability, though the best performing partition did meet this requirement above a tumor size of approximately 40 mm.

Phantom Study Evaluation of Hepatic Lesion Volume Sizing Tools

- Aims
  - Assess performance of two hepatic lesion sizing tools through anthropomorphic phantom studies
  - Estimate impact of various CT scanning conditions on hepatic lesion sizing error
- Anthropomorphic phantom
  - Customized 2-phase liver phantom

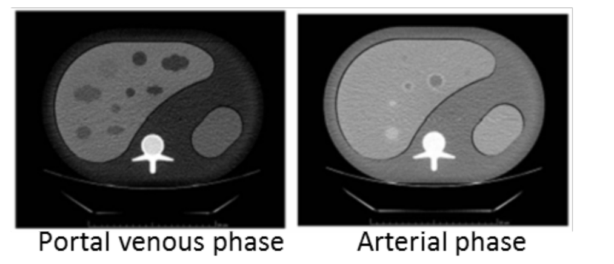


  - 19 customized synthetic lesions
    - 6-40 mm, hypo- & hyper-dense



- CT scanning of phantom
  - 2 scanners, 3 dose, 4 slice thicknesses x 2 kernels

Scanners	kVp	mAs	Pitch	Effective mAs	Rotation Time (sec)	Collimator Configuration (mm)	Slice Thickness (mm)	Overlap	Reconstruction Algorithm
GE 750HD (64 slices)	120	70 137.5 345	1.375	50 100 250	0.5	64*0.625	5 2.5 1.25 0.625	0%	Standard Soft
Siemens mCT (64 slices)	120	67.5 135 338	1.35	50 100 250	0.5 0.5 1.0	64*0.6 (32*0.6 mm detector width)	5 3.0 1.5 0.6	0%	B20f (B20s for 250 mAs) B30f (B30s for 250 mAs)



Portal venous phase — Arterial phase

- Lesion volume estimation algorithms
  - **Seg**: supervised marker-controlled watershed segmentation
  - **MF**: unsupervised matched-filter estimator (no human corrections)

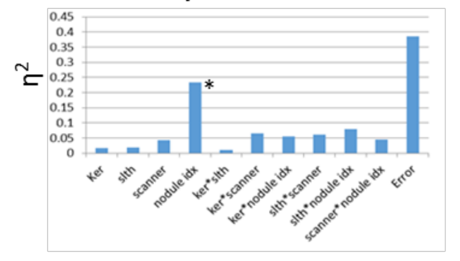
- Results: Scanner comparison
  - Data analyzed: High dose, portal venous
    GE>bias, Siemens>variance, similar RMSE

Scanner	Bias	Std. Dev.	RMSE
GE	0.16±0.20	0.16±0.11	0.25±0.20
Siemens	0..03±0.16	0.24±0.27	0.28±0.28

  - Reproducibility *Mixed-density lesions effect

Scanner	All	Solid ≥ 10mm	Solid ≥ 20mm
GE	0.55	0.35	0.27
Siemens	0.83*	0.32	0.22

  - Factor analysis *Lesion largest factor



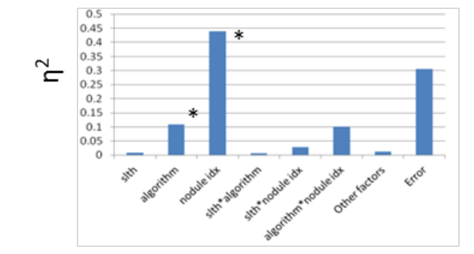
- Results: Sizing algorithm comparison
  - Data analyzed: GE, High dose, portal venous
    Seg>bias, Similar variance, Similar RMSE

Algorithm	Bias	Std. Dev.	RMSE
Seg	0.22±0.22	0.16±0.12	0.28±0.24
MF	0..001±0.27	0.15±0.13	0.23±0.23

  - Reproducibility MF more reproducible

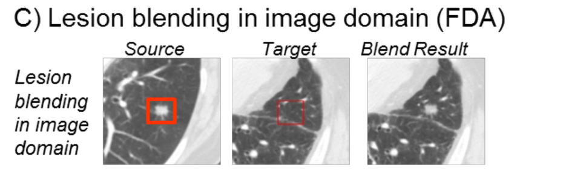
Algorithm	All	Solid ≥ 10mm	Solid ≥ 20mm
Seg	0.55	0.35	0.27
MF	0.54	0.30	0.22

- Factor analysis *Lesion & Algorithm largest factors

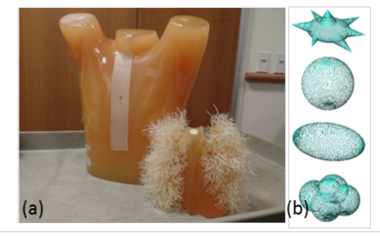


Reference Image Dataset for CT Volumetry with Known Ground Truth

- Motivation
  - Physical phantoms are useful, but often lack realism and variability
  - Clinical cases often lack ground truth
- Aims
  - Compare methods to virtually insert lung lesions of known shape and volume into clinical CT images
    - Establish technique and study influencing parameters based on physical phantoms
    - Develop a dataset of clinical CT scans with synthetic lesions for quantitative verification
    - Disseminate lesion insertion software
- Methods
  - Virtual insertion techniques
    A) Lesion mask → Projection → Recons. (Duke)
    B) Lesion mask → Recons. → Blend (Duke)
    C) Lesion blending in image domain (FDA)



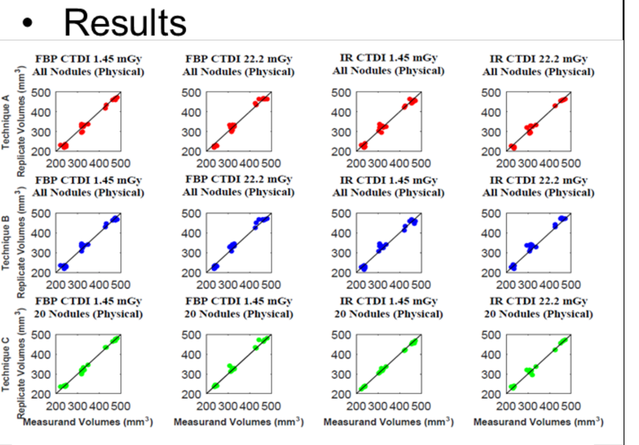
- Study Design
  - Lungman, Kyoto Phantom
  - 24 Synthetic Nodules
    - Shapes: spherical, elliptical, lobulated, spiculated
    - Nominal diameter: 8,9,10 mm



(a) An anthropomorphic chest phantom with lung insert containing realistic pulmonary vessels. (b) Virtual 3D lesion models.

  - CT scans
    - Different doses, recon methods
  - Volume estimation software
    - iNtuition (current)
    - Other commercial software (future)
  - Analysis
    - Linearity & bias

- Results



CT volumetry comparison between images of physical nodules and virtually inserted nodules across techniques A, B and C, two doses and reconstruction algorithms. Volumes were measured using iNtuition.

- Conclusion
  <3% difference for all insertion techniques between virtual and physical nodules

  $R^2 > 0.97$ for all insertion techniques

Lung Nodule Assessment in CT Screening: The "Small Nodule" Group

Purpose:

- To define evidence-based consensus standards and processes for CT imaging to allow for reproducible nodule characterization and quantification of biologically meaningful longitudinal volume changes with an acceptable range of variance across vendor platforms
- To develop standardized methods for performing repeatable volume measurements on CT images of lung nodules in the setting of ongoing lung cancer screening

Overview:

- The profile being developed addresses the accuracy and precision of quantitative CT volumetry as applied to solid lung nodules of 6-12 mm diameter. It places requirements on actors (acquisition devices, technologists, radiologists, reconstruction software, and image analysis tools) involved in activities (subject handling, image data acquisition and reconstruction, and image QA and analysis).
- The requirements are primarily focused on achieving sufficient accuracy and avoiding unnecessary variability of the tumor volume measurements.

Table 1: Modeling and simulation data that provides the basis for the profile claims

Nodule Diameter (mm)	Nodule Volume (mm³)	Coefficient of Variation (CV)
≥ 6 and < 8 mm	≥ 113 and < 268	0.29
≥ 8 and < 10 mm	≥ 268 and < 524	0.19
≥ 10 and < 12 mm	≥ 524 and < 905	0.14
> 12 mm	> 905	0.11

The QIBA Profile:

- Claim 1: Nodule Volume
  - For a nodule with diameter ≥ 6 mm and < 12 mm (volume ≥113 mm^3 and < 905 mm^3) with a measurement CV (coefficient of variation) as specified in Table 1, the following holds:
  - Claim: For a measured nodule volume of Y, the 95% confidence interval for the true nodule volume is Y ± (1.96 □Y □CV)
- Claim 2: Nodule Volume Change Between Two Time Points
  - For a nodule at time point 1 with diameter ≥ 6 mm and < 12 mm with measurement coefficients of variation CV1 and CV2 corresponding to the volume at time point 1 and time point 2 as specified in Table 1, the following holds:
  - Claim: A measured change in nodule volume of X% indicates that a true change in nodule volume has occurred if X > (2.77 x CV1 x 100), with 95% confidence. To quantify the amount of change, if Y_1 and Y_2 are the volume measurements at the two time points, then the 95% confidence interval for the true change is (Y_2-Y_1) ± 1.96 □□([Y_1 □CV1]^2 + [Y_2 □CV2]^2).
- An online calculator for investigational use in validating these claims has been proposed and is available at http://accumetra.com/NoduleCalculator.html