Quantitative Imaging Biomarker Alliance

Quantitative Imaging Biomarkers Alliance

RSNA

LUNG CANCER WORKSHOP 2013: Status Update on QIBA/FDA/FNIH Qualification

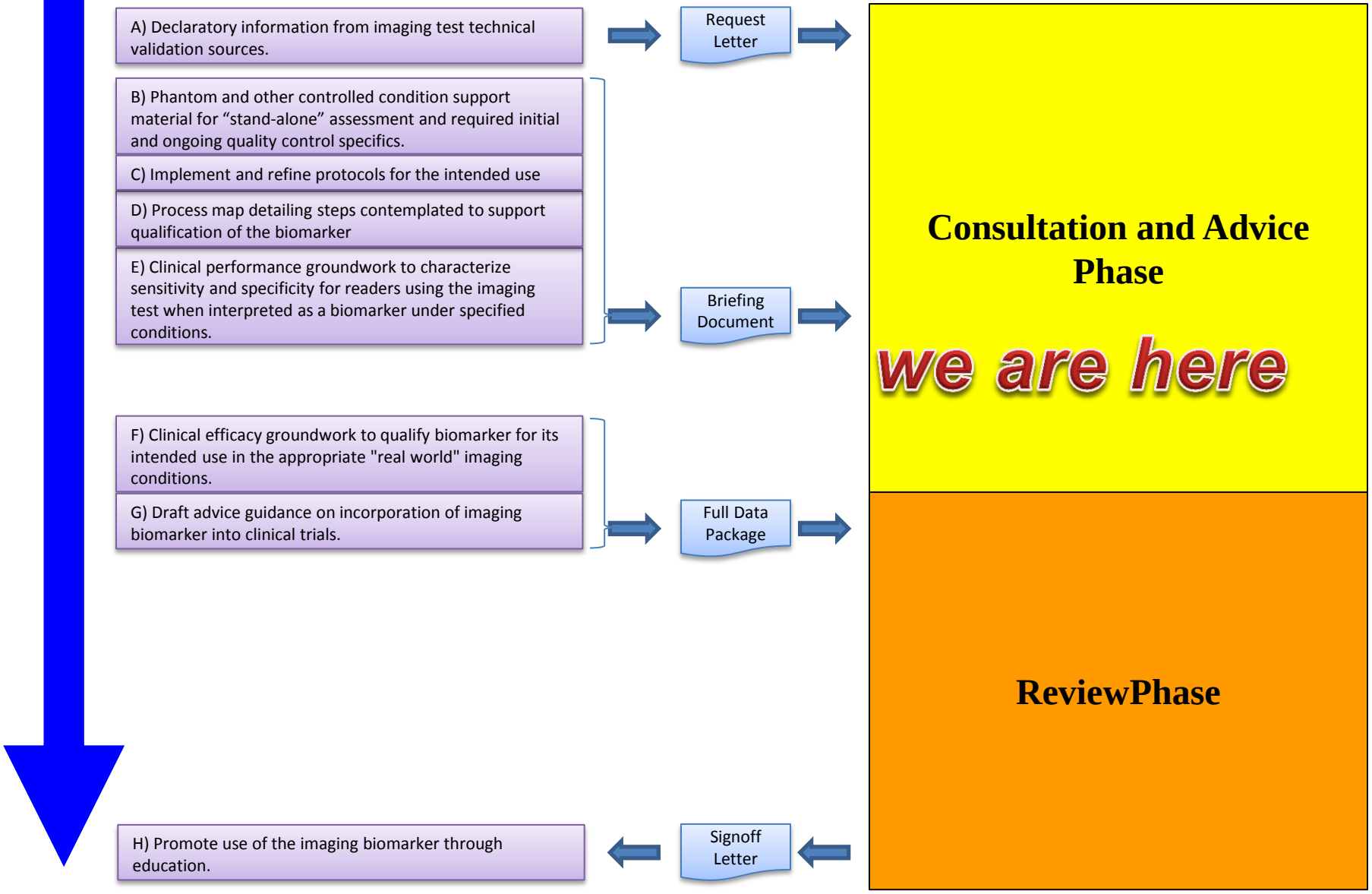
2 May 2013

Andrew J. Buckler,

BBMSC

PRINCIPAL LOGISTICAL AND FINANCIAL SUPPORT PROVIDED BY RSNA

Biomarker Qualification Process



CT Volumetry as a Biomarker

- Briefing document submitted June, 2011 followed by F2F Meeting with FDA BQRT August 31, 2011
- Submitted response September 27, 2012 addressing FDA questions and including data from:
  - Comparison of 1D, 2D and 3D nodule sizing methods by radiologists for spherical and complex nodules on thoracic CT phantom images (Petrick) (aka 1A)
  - Inter-scanner/Inter-clinic Comparison of Reader Nodule Sizing in CT Imaging of a Phantom (Fenimore) (aka 1C)
  - Comparative evaluation of multiple methods in estimating inter-method variability on nodules from CT scans of anthropomorphic thorax phantoms (Athelogou) (aka first 3A)
  - Minimum Detection Limit, Short Term Reproducibility Using Clinical Data from RIDER (McNitt-Gray) (aka, first 1B)
  - Pilot Comparison of Volumetry vs. LD on a Retrospectively Analyzed Clinical Trial Sponsored by Merck
- Received responses from agency November 20, 2012

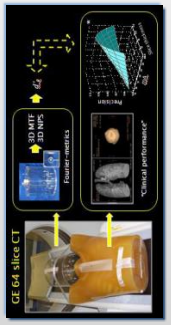
Claim Template from Metrology Workshop

- (First) Biomarker/Measurand
  - (First) Clinical Context
    - Technical Assessment:
      - Cross-sectional measurement (if specified)
        - » List Indices:
          - Bias Profile (Disaggregate indices)
          - Precision Profile
            - Test-retest Repeatability (Repeatability Coefficient)
            - Reproducibility (Intra-class Correlation Coefficient [ICC]; Concordance Correlation Coefficient [CCC], Reproducibility Coefficient [RC]):
              - Specify conditions, e.g.,
                - Measuring System variability (hardware and software)
                - Site variability
                - Operator variability (Intra- or Inter-reader)
      - Longitudinal change measurement (if specified)
        - » List Indices: (as above, including sub-parts)
    - Clinical Assessment:
      - Association of biomarker for predicting patient outcome
  - (Next clinical context, if any) ...
- (Next biomarker/measurand – if any) ...

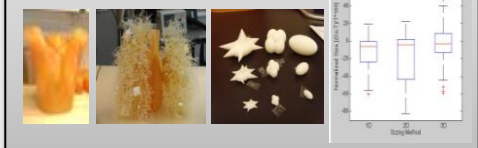
Process Map: Series of Directed and Related Studies that build on each other

PROFILE Authoring and Testing

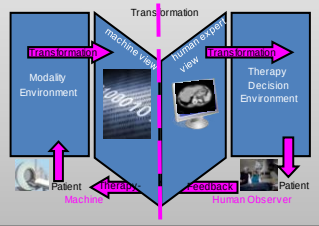
Explore figures-of-merit and QC procedures



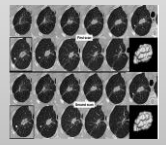
Intra- and inter-reader variability (1A)



Inter-analysis technique (algorithm) variability (3A)

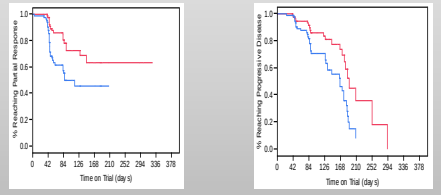


Minimum detectable biological change (1B)



5 readers, 3 reads each

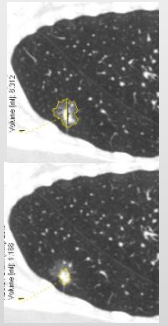
Correlation with clinical endpoints and outcomes (3B)



Inter-scanner model, and -site variability (1C)

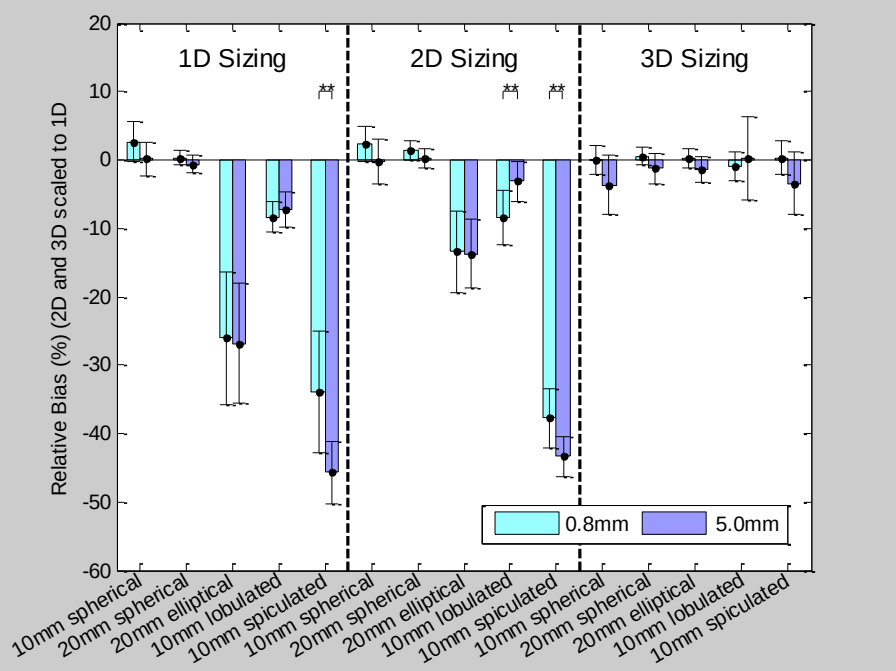
Performance-based branch / compliance procedure

Extend to other Lesion characteristics

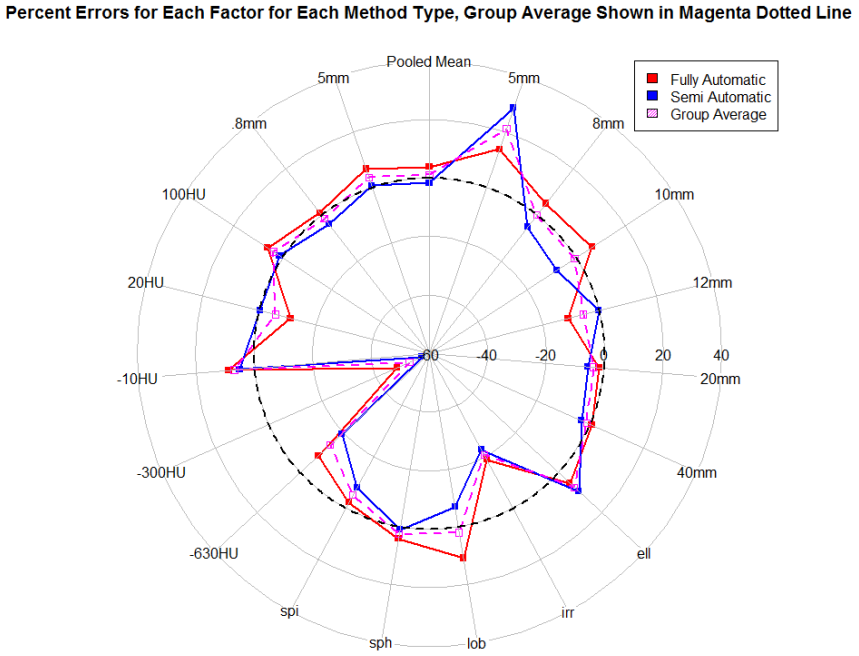


Solid Masses in Lung: Cross-sectional Measurement: Relative Bias

Range of scanner settings, 1D/2D/3D, single method (1A)

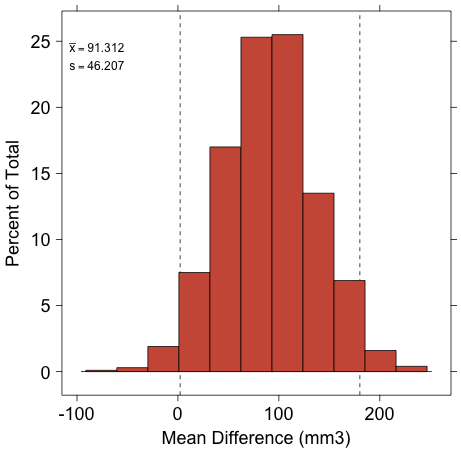


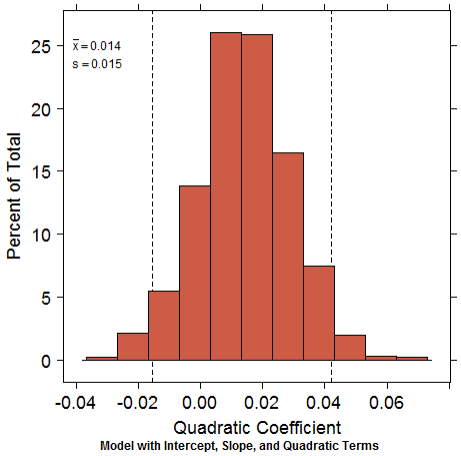
(same) Range of scanner settings, multiple segmentation methods (3A)

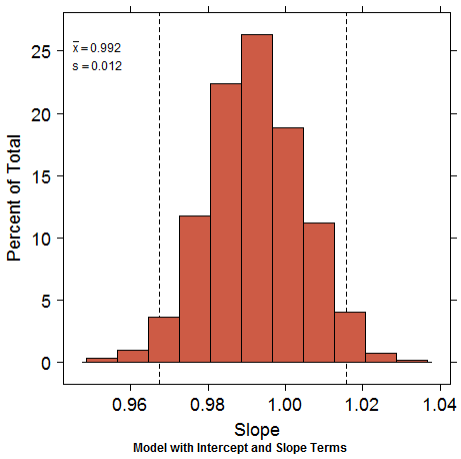


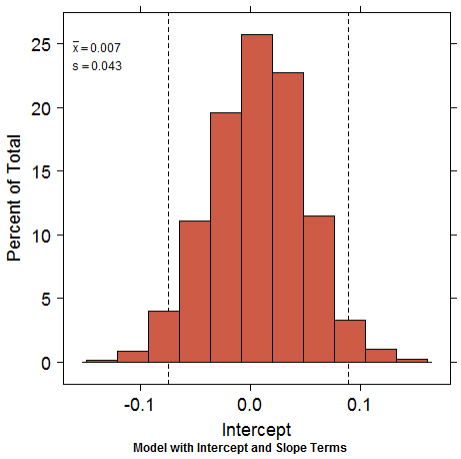
Unlike 1D and 2D, volume 3D is essentially an unbiased estimator, even with irregular shapes. Moreover, this result is robust across multiple segmentation methods at various levels of automation.

Cross-sectional Measurement: Absolute Bias and Linearity









Percentile	Mean Difference (log(mm³))	Quadratic Coefficient	Intercept	Slope	CCC	Slope Only
2.5%	0.002	-0.015	-0.075	0.967	0.973	0.991
97.5%	0.030	0.042	0.089	1.016	0.984	0.998

The following may be concluded based on this analysis:

- True mean bias is between 45mm^3 and 137mm^3 on measurements with geometric mean of 1023 mm^3 (4%-14%). The interval is near but does not contain zero, with 95% confidence.
- The magnitude of the quadratic coefficients is essentially zero. Hence, if a quadratic effect is truly present, its impact is negligible.
- For the model with a slope and intercept, the non parametric 95% confidence interval for the intercept contains zero and that for the slope contains 1, establishing linearity over the range 4-34,389mm^3 (up to about 34cc). The distribution of the CCC does not contain 1, but is not practically different from 1.

Taken as a whole, this evidence establishes the basic characteristics needed for use of CT as an unbiased estimator of tumor size useful as a measure of growth in clinical trial settings.

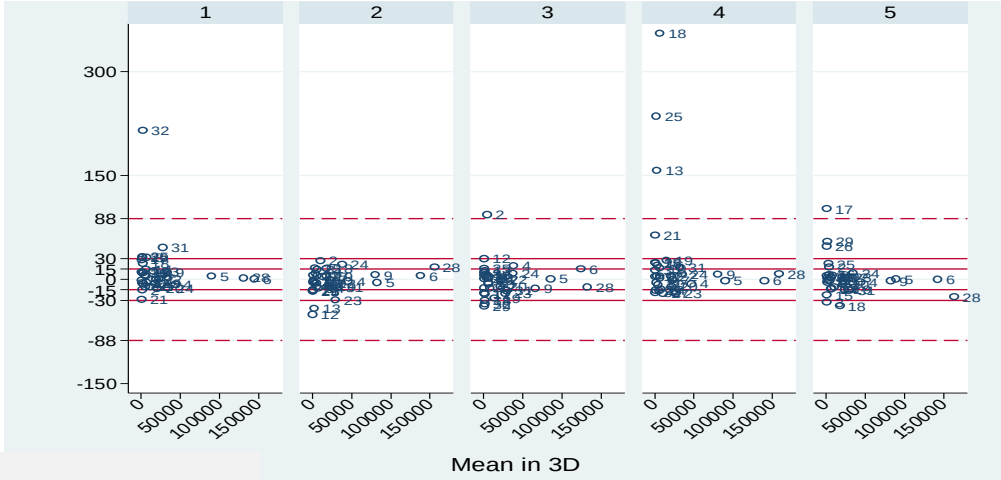
Precision Profile: Test-retest Repeatability

Hypothesis

That the minimal detectable change in tumor size - using measured tumor volumes made by radiologists on thin section CT images - will be smaller when using a side by side ("clinical trial workflow") review setting than when using an independent review setting (previous study)

Results

	Independent Reads [95% CI]	Locked, Sequential Reads [95% CI]
%Volume Change	[-4.0%, 55%]	[-2.2, 17%]
% Diameter Change	[-0.3%, 12%]	[0.3%, 4.7%]



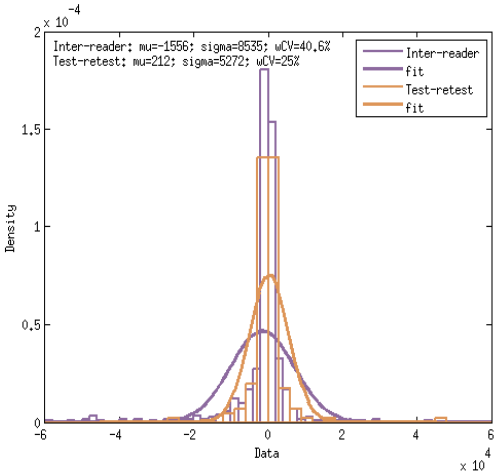
Discussion

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

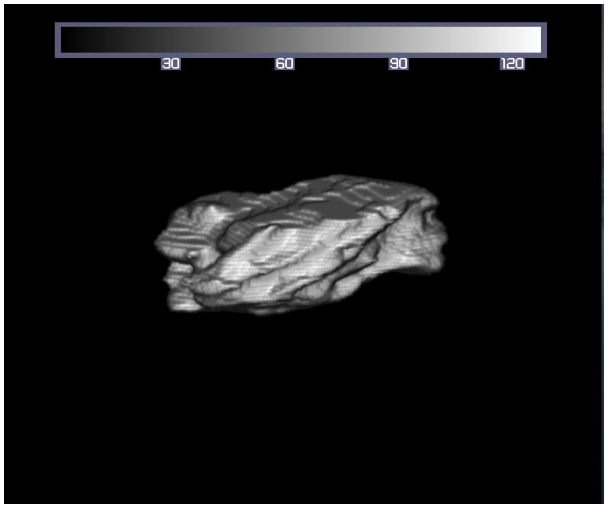
Test-retest Repeatability: Extended Analyses, teasing apart reader effect

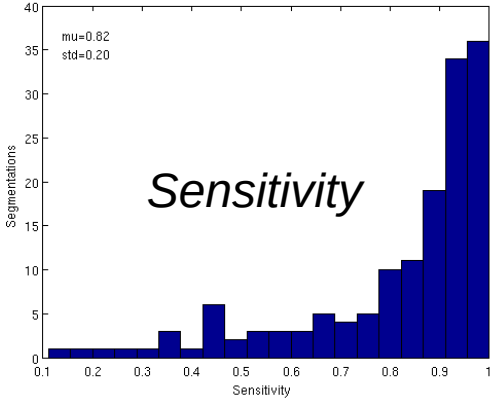
- Findings from Bland-Altman analysis: for Test-retest: wCV 29%, Inter-reader: wCV 59%. The results are not uniform at the lower end of the reported range versus the upper end.
- Test-retest, and inter-reader CCC are .99, and .93, respectively, using Lin's method. Corresponding standard deviations of .12 log(mm^3) and .28 log(mm^3) which agree with the Bland-Altman analysis.
- Findings from probability density function of error analysis: Standard deviation of error was calculated to be on the order of +/-5272, and 8535 mm^3 for test-retest repeatability, and inter-reader reproducibility respectively on TV data. These numbers are comparable to those produced by the Bland-Altman estimates and in line with the mixed effects model result.
  - wCV_{TV} is 25% for test-retest, and 41% for Inter-reader considering mean TV=21002 mm^3, which is lower than that which is concluded from the other analyses.

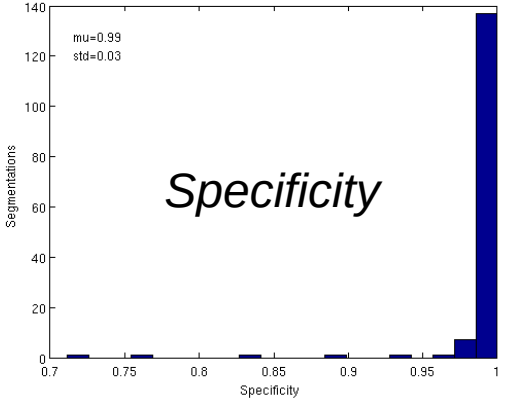
As a grand summary, it may be concluded that the test-retest wCV lies between 25 and 29% while the inter-reader wCV lies between 41 and 59%.

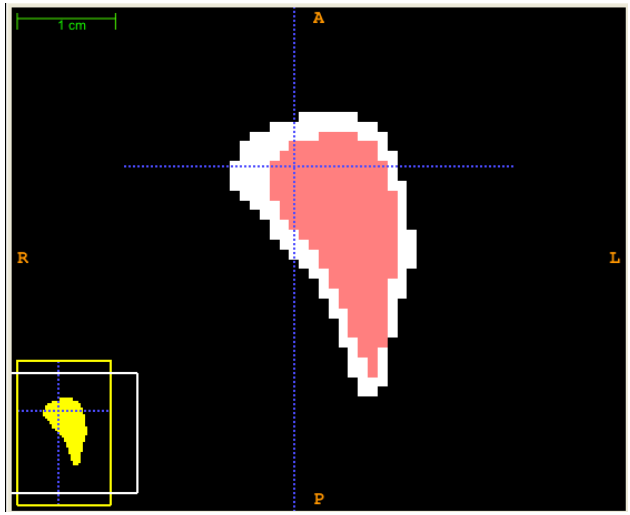


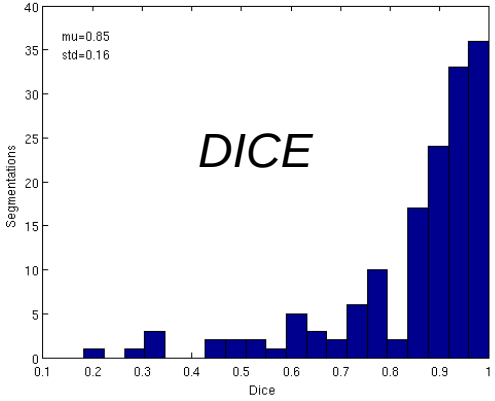
Test-retest Repeatability: Extended Analyses, Overlap Metrics

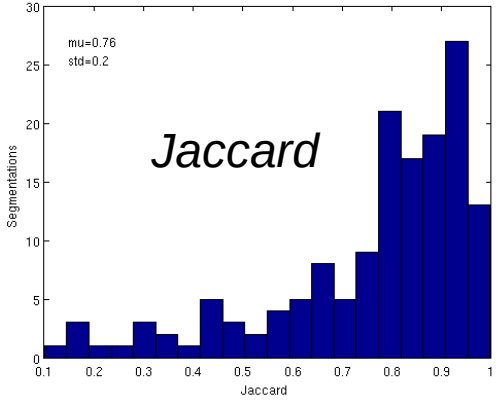












Label map files from second 1B study

(steps only needed to accommodate 1B)

- Segmentation Objects: 1.2…UID.dcm
- Extract tags such as SUBJID, readerID, Acqrep
- Rename to Seg_SUBJID_0_ACQREP_rdrID.dcm → Convert from dcm to nii format
- Extract tags such as pixel spacing, matrix
- Calculate volumes

Primary Analysis

- Prepare input files for statistical modules
- Bland Altman, CCC
  Linear Mixed Effect
  Probability Density Function of Error

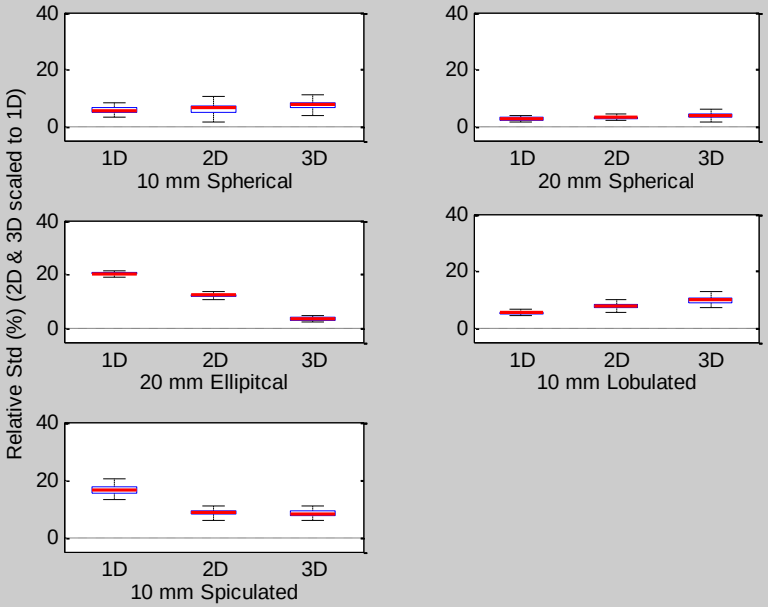
Secondary Analysis

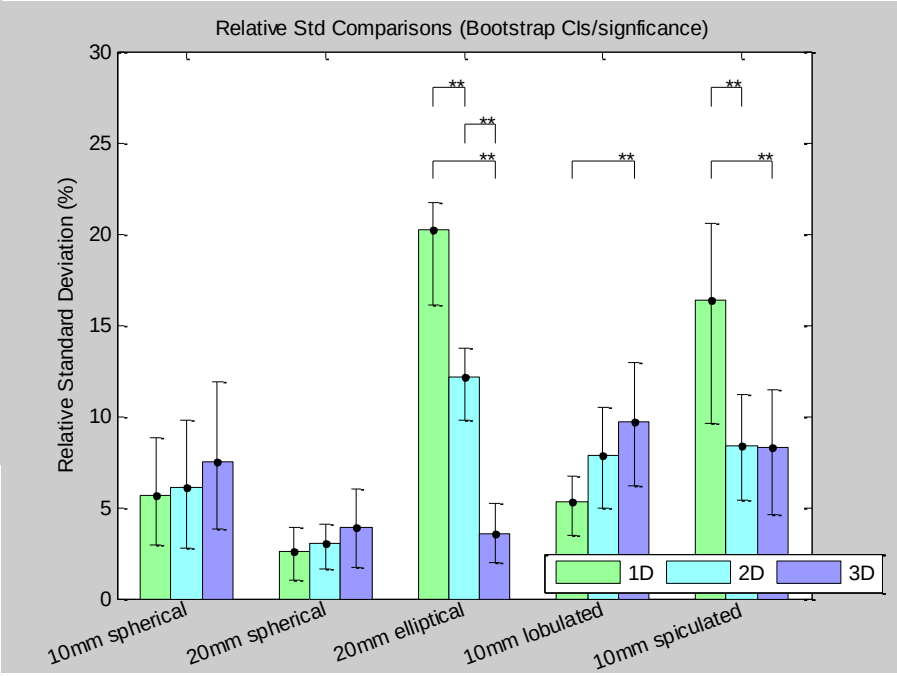
- Register all seg_SUBJID_0_ACQREP_rdrID.nii with respect to one of them within same acqrep
- For each SUBJID and for each ACREP
  STAPLE → Weights.nii
  Binarize → Weights-bin.nii
  Pair-wise comparison between seg_SUBJID_0_ACREP_rdrID.nii-r.nii and Weights-bin.nii
- Create input files for analysis of metrics: Intersection, Union, Jaccard, DICE

Tertiary Analysis

- Register all seg_SUBJID_0_ACQREP_rdrID.nii with respect to one of them across different acqrep

Precision Profile: Reproducibility: Across Scanner Settings and Shapes

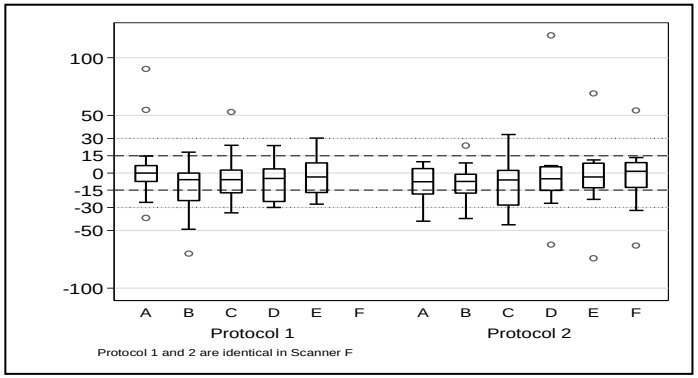


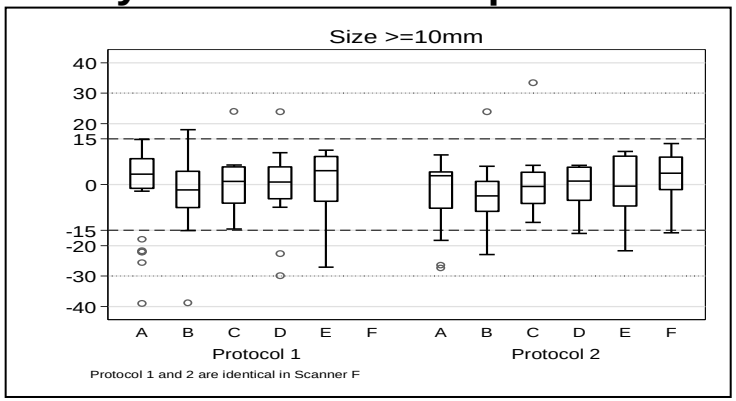


Precision Profile: Reproducibility: Across Scanner Models and Sites

Aim is to characterize accuracy and precision in reader measurements of volumes of six phantom nodules in CT imagery collected on six scanners.

Summary Data: Percent Relative Bias in volume by Scanner and CT protocol





Conclusions:

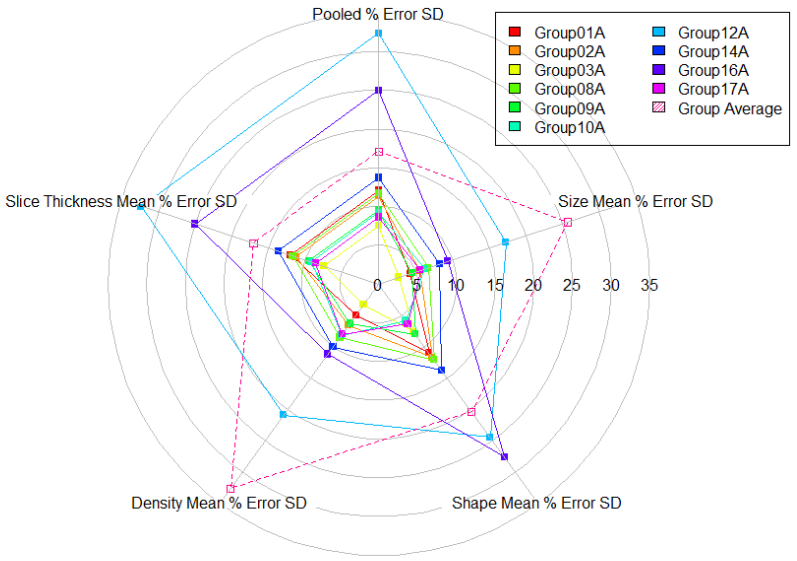
- Relative bias in pooling the 6 nodules is within a tolerance of 15%.
- In an equivalence t-test applied to each of the 6 nodules, scanner equivalence is found only for the larger synthetic lesions (10 mm and 20 mm). This finding confirms the lesion sizing guidance (10 mm and up) in QIBA CT imaging profile.
- Equivalence of the protocols supports the imaging protocol as used by ACRIN Trial 6678.

Precision Profile: Reproducibility: Across Algorithms and Methods

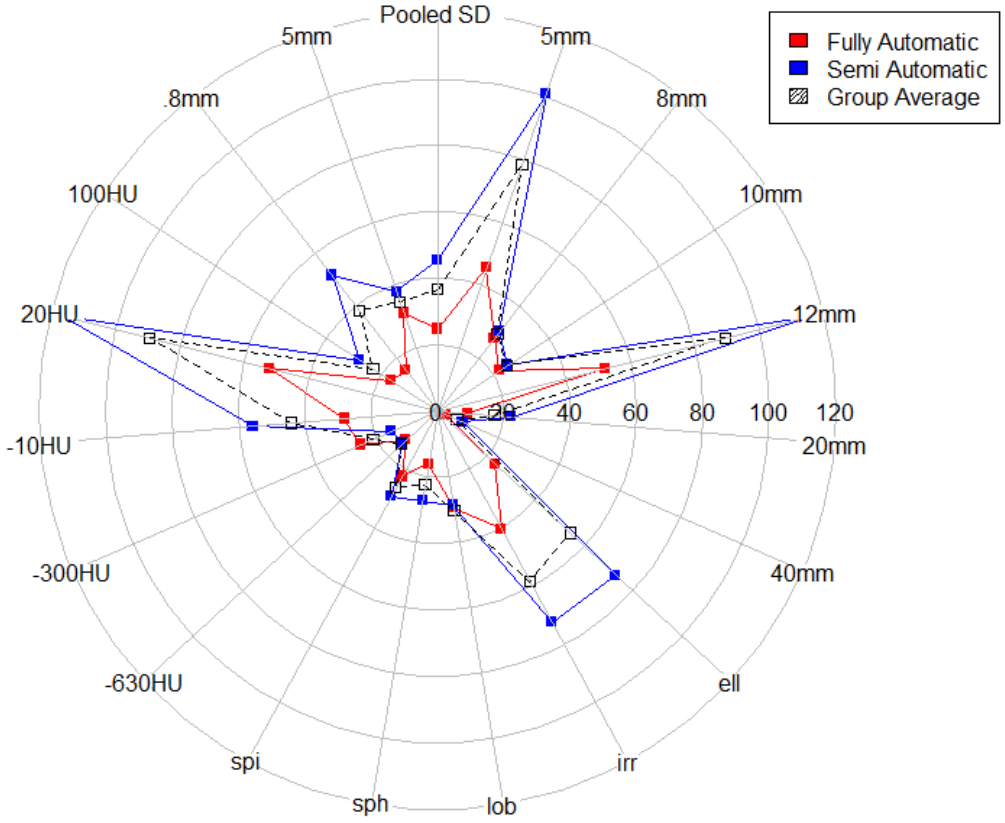
Challenge Definition: estimate absolute volumes in phantom data Explicitly indicate descriptive statistics: **bias, variance.**

Null hypothesis: analysis software model does not have a significant effect on the bias and variance.

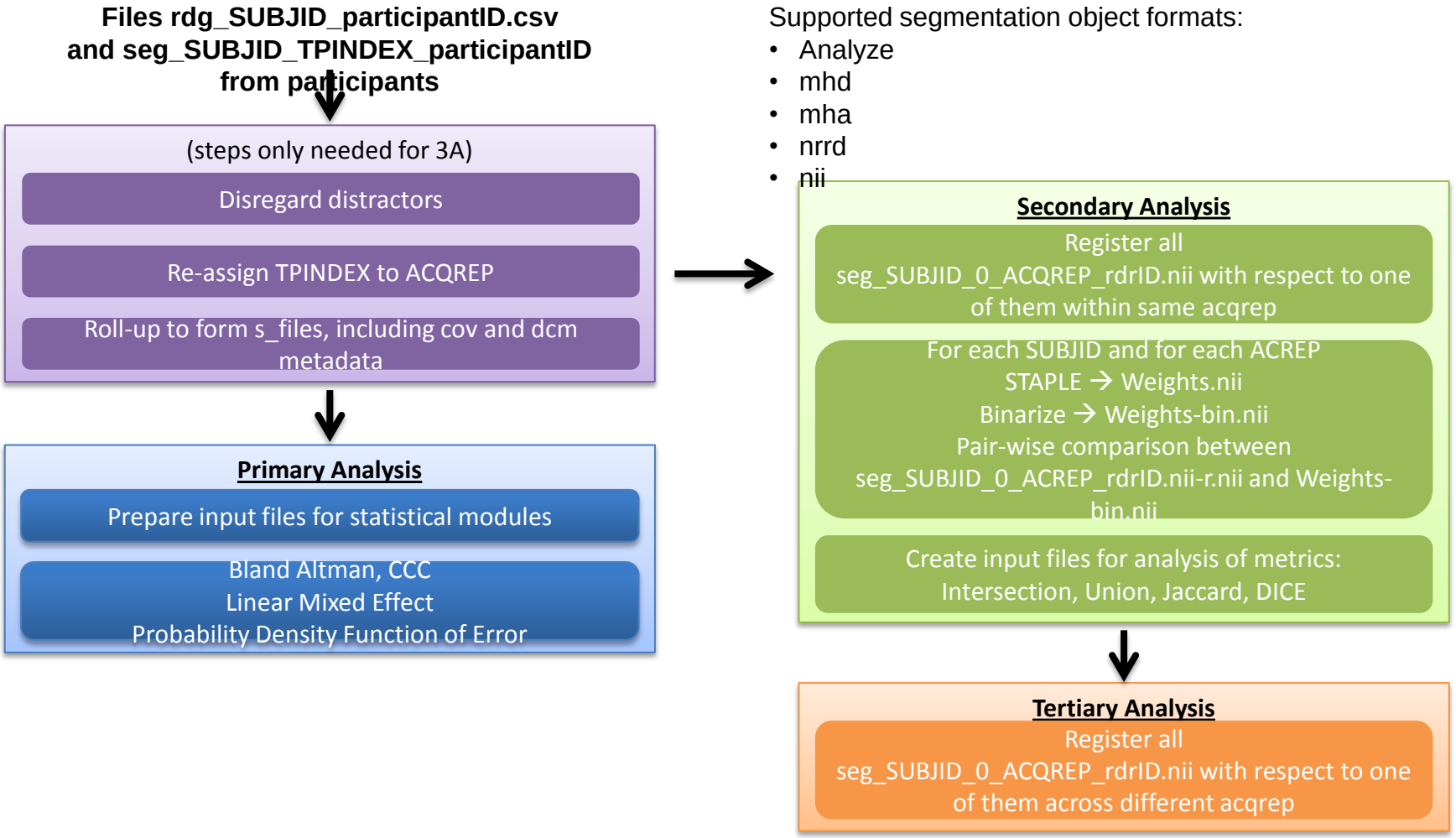
Percent Error SDs for Each Factor, Group Average Shown in Dotted Line



SD Percent Errors for Each Factor by Method, Group Average shown in a Dotted Line



(next challenge, currently accepting applications)



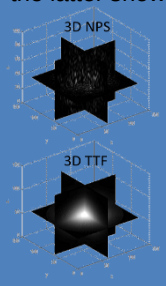
Precision Profile: Reproducibility: Across Reconstruction Techniques and Dose

Aims

Develop and evaluate a metric (estimability index, e') capable of modeling/ predicting the performance of chest CT volume quantification.

System resolution and noise information

TTF and NPS can be measured from either ACR or Mercury phantom, with the latter showing more flexibility in mimicking various patient sizes.



Noise power spectrum (NPS) described the noise component in the frequency domain. It is measured at multiple noise levels to capture the possible shift of frequency components as observed in iterative reconstructions.

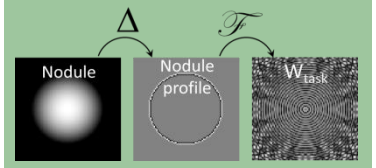
Task transfer function (TTF) is an extension of the concept of modulation transfer function (MTF) to accommodate the nonlinearity of iterative reconstruction within a narrow, linearizable operating band (i.e., specific object contrast and noise levels). It is acquired at various contrast and noise levels to fully characterize the whole operating space.

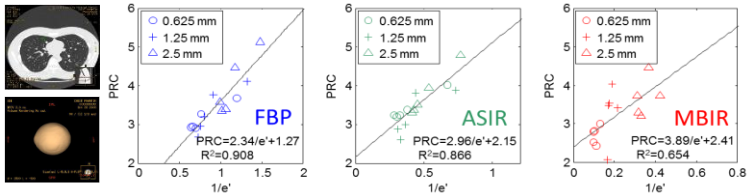
Quantification software info.

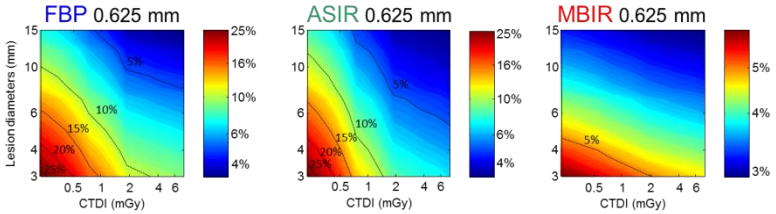
Internal noise (N_i) reflects the inconsistency of the quantification software due to differences in the placement of random seeds. It is measured from nine locations of a nodule.

Nodule information

Task function (W_{task}) describes the size, shape, and contrast of the nodule to be quantified in the frequency domain.







Current Status

Standardize the calculation of e' and recommend guidelines for compliance of quantification techniques

Precision Profile: Reproducibility: Across Lesion Density Characteristics

Aim

Extend characterization of nodule measurement performance to the part-solid case in a phantom study. Primary endpoints include bias and variability relative to known nodule volume, with covariates including dose, slice thickness, nodule shape, size, and mean CT density.

Methods

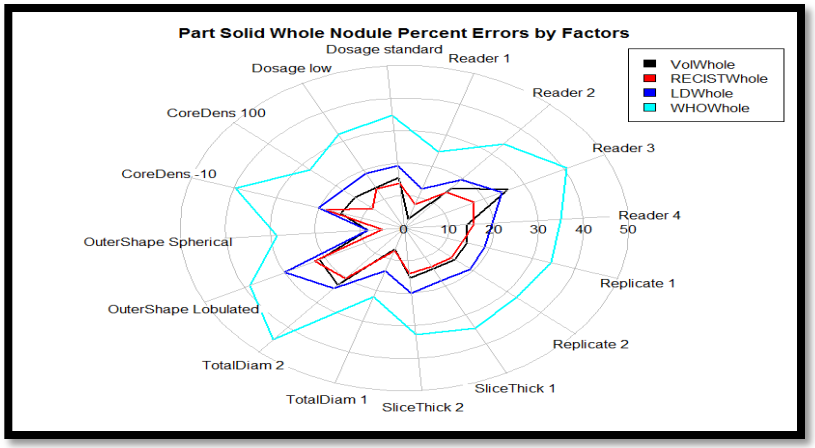
Part Solid and Solid Nodules
Shape, Size (Whole/Core), Density (Outer,Core)
Spherical, 10mm/5mm, -630 HU/-10HU
Spherical, 10mm/5mm, -630 HU/100HU
Lobular, 10mm/5mm, -630 HU/-10HU
Lobular, 10mm/5mm, -630 HU/100HU
Spherical, 20mm/10mm, -630 HU/-10HU
Spherical, 20mm/10mm, -630 HU/100HU
Lobular, 20mm/10mm, -630 HU/-10HU
Lobular, 20mm/10mm, -630 HU/100HU
Spherical, 12mm/NA, 100HU/NA
Spherical, 10mm/NA, 100HU/NA
Spherical, 8mm/NA, 100HU/NA
Spherical, 5mm/NA, 100HU/NA

Acquisition Protocol
QIBA Profile. Computed Tomography: Change Measurements in the Volumes of Solid Tumors, Version 2.0
Siemens Sensation 64
Slice Thickness: 1mm, 2mm
Dose: 40mAs, 200mAs

Reading Protocol
Four Radiologists, 2 reads each
Semi-automated (SA), seed-based, INTIO research software
Data output (ALL based on the 3D segmentation):
Nodule Volume
SA-RECIST: Axial plane diameter
SA-LD: Longest diameter in an orthogonal plane
SA-WHO: Area based on WHO criteria

Significant Preliminary Results – Whole Nodule Measurements

- Absolute Bias of Whole PS Volume > SA-RECIST
  - PS Volume Absolute Bias = 20.38
  - SA-RECIST Absolute Bias = 15.64
  - SA-RECIST taken from volume segmentation
  - Spherical Nodule bias
- Significant Covariates of Volume and SA-RECIST
  - Nodule Diameter
  - Nodule Outer Shape
  - Reader (R)
- Significant Covariates of SA-RECIST only
  - Nodule Core Density
- ICC (95% CI)
  - Whole Volume Measures – Reader Mean, 0.955 (0.83, 1.00)
  - Absolute bias from Volume – Reader Average, 0.302 (0.00, 0.79)
- No significant effect of dose



Lung Cancer: Longitudinal Measurement

- If the linearity assumption is shown to be reasonable for the range of plausible values of X, then one can use a simple error propagation formula to estimate the precision of the estimated change from the cross-sectional estimate of precision. Let s(Y) denote the estimated precision of Y at a single time point. s(Y) is often expressed as the standard deviation of Y but other measures of precision are also common. If s(Y) is a constant value not related to the value of X, then an upper bound (assuming a positive correlation) on the precision of an estimate of the change between time t=*0* and t=*t* is given by

$$s(Y_0\text{-}Y_t) = \mathrm{sqrt}(2\times[s(Y)]^2). \qquad [1]$$

- For example, let wSD be the within-subject standard deviation of a QIB algorithm measuring nodule volume at a single time point. Suppose wSD is 15. Then the estimated within-subject standard deviation of the change in nodule volume, wSD_Δ, is 21 [15]. If, on the other hand, s(Y) changes in magnitude with, say, the true size of the lesion, X, then a reasonable upper bound on $s(Y_0\text{-}Y_t)$ is given by:

$$s(Y_0\text{-}Y_t) = \mathrm{sqrt}([s(Y_0)]^2+ [s(Y_t]^2), \qquad [2]$$

where $s(Y_0)$ and $s(Y_t)$ are the precision estimates of the nodule volume at baseline and time *t*, respectively.

- The estimates of uncertainty in change measurements in equations 1 and 2 do not take into account the within-subject correlation. The within-subject correlation is the correlation in the measurements at the two time points due to the fact that it is the same lesion in the same patient being measured at two time points. The simple formulae in equations 1 and 2 provide only upper bounds on the precision. A more appropriate formula is

$$s(Y_0\text{-}Y_t) = \mathrm{sqrt}([s(Y_0)]^2+ [s(Y_t]^2 - 2\times r\times s[Y_0)] \times s[Y_t)]). \qquad [3]$$

	Lesion Size (mm) [*RDC* (mm³), *RDC %CV*]		
Algorithm	8-10 (n=36)	20 (n=44)	40 (n=10)
1	260, 50%	820, 19%	1830, 5%
2	290, 62%	1100, 28%	8100, 26%
3	740, 110%	1980, 40%	2050, 6%
4	160, 28%	570, 13%	1330, 4%

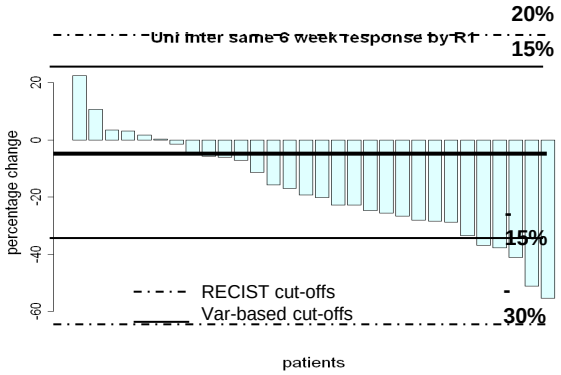
Clinical Assessment (colorectal cancer)

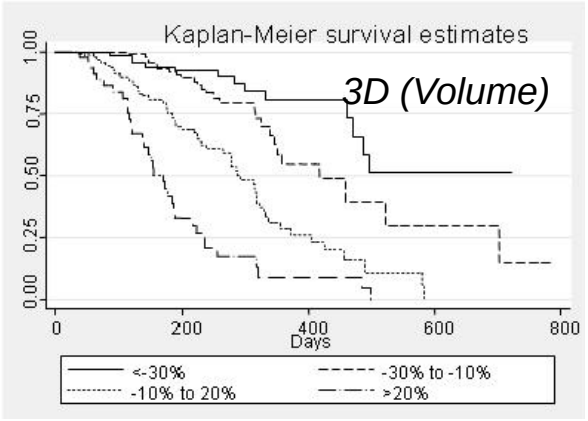
Specific Aims

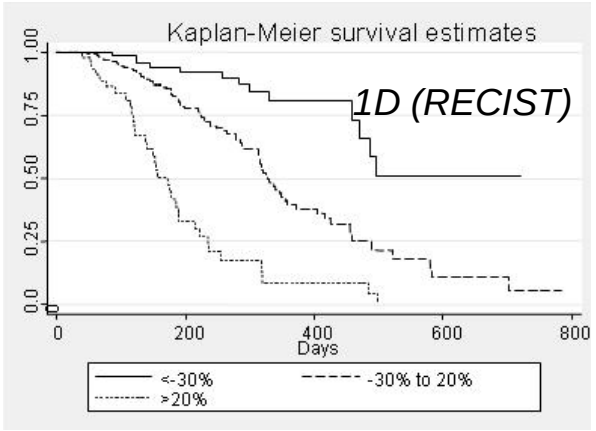
1. To explore variability in measuring change in tumor volume (uni- and bi-dimensional as well)
2. To correlate responses assessed by the volumetric, uni-dimensional and bi-dimensional measurement techniques with patient survival

Materials and Methods

- We used an image dataset of 560 patients enrolled in a multicenter Phase II / III clinical trial of advanced colorectal cancer and treated with a targeted therapy.
- Target lesions were measured on baseline and follow-up scan time-points using in-house lesion segmentation algorithms developed for solid tumors.







Our preliminary results indicate that the minor change category of -30% - -10% may correlate with longer survival compared to the -10 – 20% group.

Clinical Assessment (ACRIN 6678)

To evaluate in an exploratory analysis changes in tumor volume during chemotherapy by multislice CT:

3.3.3 Can changes in tumor volume be assessed by multi-detector CT early during the course of chemotherapy?

3.3.4 Are tumor volumetric changes correlated with patient outcomes?

3.3.5 Can one develop parameters that combine metabolic and volumetric data and do these parameters allow a better prediction of patient outcome than metabolic changes alone?

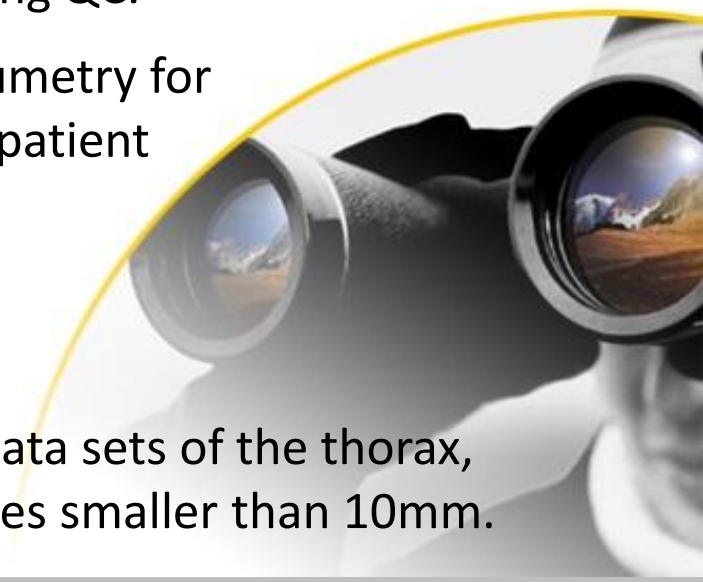
Next Briefing Document Update (with which we will request the next F2F with BQRT)

- Validation of Volumetric CT as a Biomarker for Predicting Patient Survival (Schwartz) (aka 3B)
- Development of Assessment and Predictive Metrics for Quantitative Imaging in Chest CT (Samei) (aka Duke)
- Quantifying Variability in Measurement of Pulmonary Nodule (Solid, Part-Solid and Ground Glass) Volume (Garg) (aka Colorado)
- Assessing Measurement Variability of Lung Lesions in Patient Data Sets: Variability Under Clinical Workflow Conditions (McNitt-Gray and Clunie) (aka second 1B)
- Inter-algorithm Performance Investigation on Clinical Data (Athelogou) (aka second 3A)
- (pending) Cohorts by Nishino (80 NSCLC pts) and Hoffe (14 pancreatic pts)
- Abigail?

Next up: 3 year deliverables as presented to the RSNA board for CT volumetry

Advanced disease:

1. Develop Profiles for CT volumetry of hepatic masses and lymphatic metastases.
2. Characterize comparative algorithm performance for patient data sets in the thorax and abdomen.
3. Apply predictive metrics for CT volumetry in a calibration and quality control program for both compliance testing and ongoing QC.
4. Conduct additional validation studies of CT volumetry for FDA qualification as a biomarker for predicting patient survival in an expanded range of indications.



Screening:

1. Using data from phantom studies and patient data sets of the thorax, develop a Profile for CT volumetry of lung masses smaller than 10mm.