

QIBA CT Volumetrics Group 1B: (Patient Image Datasets)

Update
April 19, 2011

Experiments to Explore First:

1. What level of **bias and variance** can be achieved in measuring tumor volumes in patient datasets?
2. What is the **minimum detectable level of change** that can be achieved when measuring tumors in **patient datasets** under a “No Change” condition?

Agenda

- Review
 - Study Aims, Designs, Methods
- Analyses
 - Pooled
 - Subjective 1 (Mike O'Neal)
 - Data Driven (over Grace's objections)
 - Subjective 2 (Mike McNitt-Gray)
 - Min. Detectable Change
- Next Steps

Experiment 2 – What is the min detectable level of change in patient datasets under a “No Change” condition?

1. Specific Aims

- (a) For patient datasets acquired over a very short time interval (presumably the “no change” condition) investigate variance of both readers and algorithm-assisted readers in measuring **change** in volume, diameter and bi-directional diameters of lesions (here, the expected value of the **change** should be zero)
 - (b) Investigate several change metrics such as:
 - (a) Absolute value of change
 - (b) fractional change in volume/diameter
 - (c) Investigate inter-observer variability in each task
 - (d) Investigate Intra-observer variability in each task
- (NOTE: again, observer should be interpreted broadly – as reader measuring manually for diameters as well as algorithm-assisted reader measuring contours).

Methods

- RIDER – MSK Coffee Break Experiment (No Change Condition)
 - 32 NSCLC patients
 - Imaged twice on the same scanner w/in 15 minutes
 - Thin section (1.25 mm) images
 - Selected only one lesion per patient ->32 lesions.

Methods

- Multiple Markings
- Manual linear measurements (Single Longest Diameter on one image)
- Separate Manual 2 Bi-directional Diameters (Longest Diameter and Diameter perp.)
 - Single Longest Diameter is also retained for Comparison
- Separate Algorithm Assisted volume (Reader contours entire boundary of lesion).
 - Also calculate Single Longest diameter in a given image as well as perp diameter

Methods

- 5 readers
- Read each case:
 - Scan 1, Scan 2, repeat read of Scan 1 (to assess intra reader variation)
- Order is randomized by patient, scan, measurement type

Methods

1. METHODS and MATERIALS
 - To expedite lesion identification, Lesions are pre-identified and approximate locations are provided to readers.
 - This will be done using proprietary software at RadPharm.
 - Lesions were pre-identified by placing an ellipse on the 2nd or 3rd slice through the lesion (eliminating slice selection bias by the reader).
 - Reading permissions for the linear measurement application were set so that each reader can see these annotations, but no one else's.
 - Readers can move quickly to identify the lesion to performing marking task

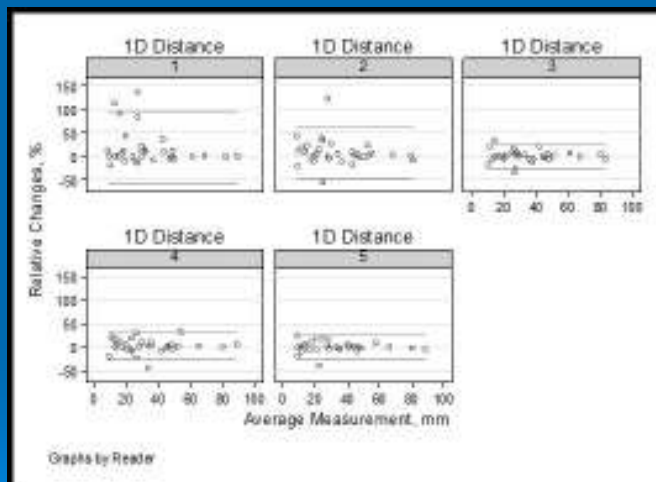
Methods

- Analyses
 - Estimate variance measured values for
 - Diameter (from Manual, Bidimensional and Vol)
 - Product of Diameters (Manual and Vol)
 - Volume
 - Estimate inter-reader variability
 - Intra-reader variability from those cases repeated by readers

Results

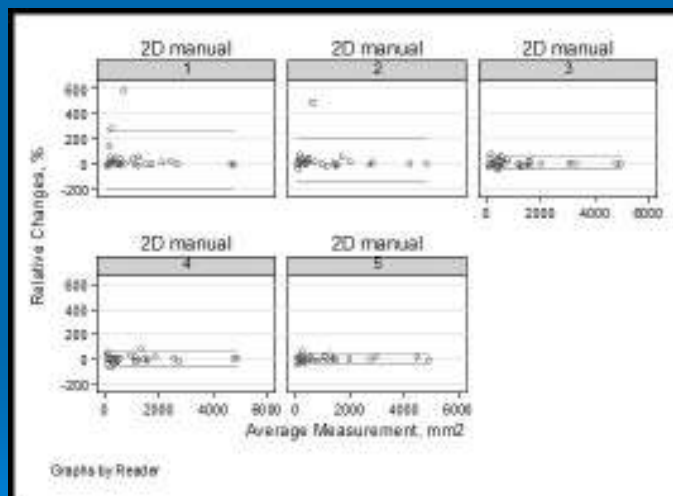
- Pooled Analysis
 - across all readers and all lesions
- Percent Difference between scans 1 and 2
 - Mean (SD)
 - 1D: 5.84 (23.83)
 - 2D: 15.22 (68.45)
 - 3D: 24.99 (117.88)

Results



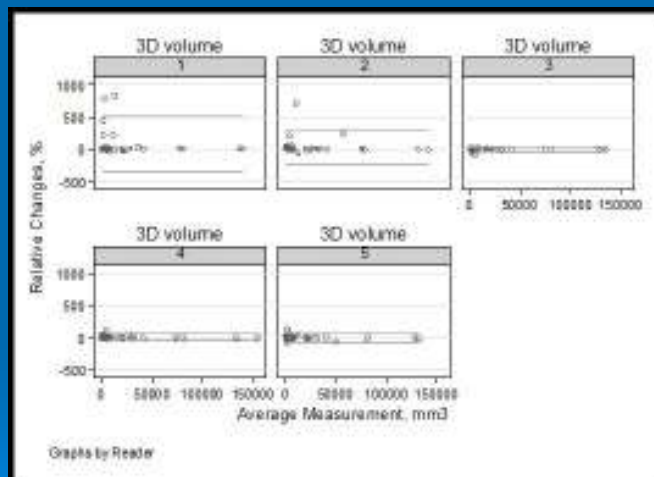
Heterogeneity across Readers and Lesions (especially small ones)

Results



Heterogeneity across Readers and Lesions (especially small ones)

Results



Heterogeneity across Readers and Lesions (especially small ones)

Subgroup Analyses – 1

- Mike O'Neal rated RECIST measurable or NO
- 28 cases were measurable, 4 NO

Method	1D	2D	3D
RECIST measurable	11.97 (21.77)	25.35 (68.92)	32.40 (101.22)
NOT	18.93 (11.79)	33.74 (21.93)	77.42 (180.43)

Subgroup Analyses – 2

- MMG rated Difficult/Moderate/Easy
- Percent Difference between scans

Method	1D	2D	3D
Easy	0.99 (.06)	2.58 (15.61)	2.85 (13.53)
Moderate	9.42 (28.57)	17.99 (61.79)	42.82 (107.49)
Difficult	8.52 (29.54)	24.40 (91.75)	36.76 (159.12)

Minimum Detectable Change

	1D [95% CI]*	2D [95% CI]*	3D [95% CI]*
All (N=32)	[-0.3%, 12%]	[1%, 29%]	[-4%, 55%]
Measurable (N=28)	[-1%, 14%]	[-1%, 32%]	[-1%, 41%]
Un-measurable (N=4)	[-8%, 13%]	[-4%, 30%]	[-31, 158%]

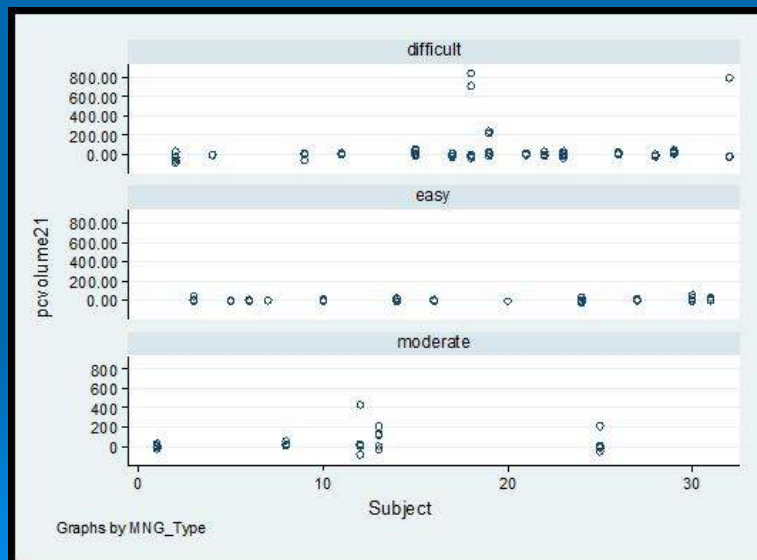
*: mixed effect model were used, where readers were random effects.

Minimum Detectable Change

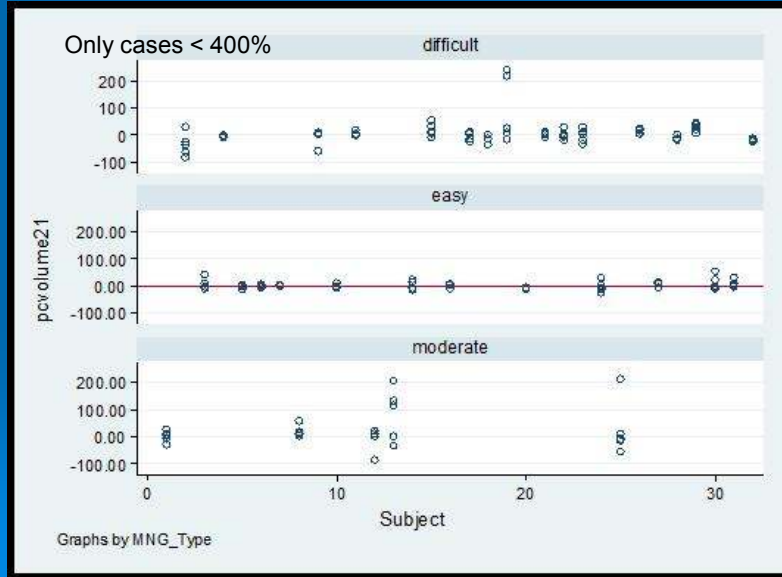
	1D [95% CI]*	2D [95% CI]*	3D [95% CI]*
All (N=32)	[-0.3%, 12%]	[1%, 29%]	[-4%, 55%]
2 nd subjective assessment – EASY (N=12)	[-1%, 3%]	[-2%, 7%]	[-1, 7%]
Moderate (N=5)	[-4, 23%]	[-17, 53%]	[-4, 89%]
Difficult (N=15)	[-1%, 18%]	[2, 47%]	[-10, 83%]

*: mixed effect model were used, where readers were random effects.

% Change in Volume by Type



% Change in Volume by Type



Next Steps

- Further Analyses?
- Present to QIBA volCT group
- Manuscript
 - One manuscript on these results
 - One editorial on implications for QIBA/Clinical Trials, etc.?