2014 Report from the CT Volumetry Biomarker Committee of the Quantitative Imaging Biomarkers Alliance

Gregory Goldmacher, MD, PhD; Samuel Armato III, PhD; Maria Athelogou, PhD; Andrew Buckler, MS; James Mulshine, MD; Nicholas Petrick, PhD; Berkman Sahiner, PhD; Ehsan Samei, PhD; Binsheng Zhao, DSc; Lawrence Schwartz, MD

Background And Previous Work

Establish 3D CT Volumetry as a Surrogate Endpoint Biomarker for Tumor Response

Specific Aims

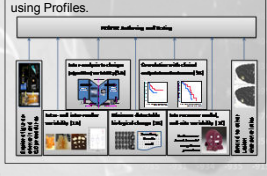
To develop methods and processes for accurate and reproducible measurements of anatomic structures and masses.

Stakeholders

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

Approach

Process map for qualifying volumetric image analysis using computed tomography, showing relationship to standardization of the biomarker's measurement and interpretation using Profiles.



Where We Go From Here

- Advanced Disease:
  1. Complete our project in lung cancer by developing methods for quantifying its metastases to the liver and lymphatic system; and
  2. Expand these results to a broader set of indications that also produce solid metastatic lesions in the liver and the lymphatic system. These tissues are common sites of disease in colorectal, esophageal, renal, breast, stomach, pancreatic, melanoma, etc.
- Lung Nodule Assessment in CT Screening

Federal Funding Acknowledgements

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.

Various QIBA projects and activities have been supported in part with Federal funds from the National Institute of Standards and Technology, Department of Commerce, under Cooperative Agreement No. 70NANB10H223 (which partially funds the QI-Bench program).

What We've Accomplished

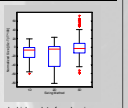
- Completed and reported charter studies (see below).
- Expanded studies to characterize performance under an increasing range of settings.
- Written, refined, and released Profile through Public Comment and Field Test processes.

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).

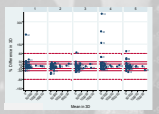
Summary

- 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



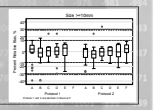
Discussion

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

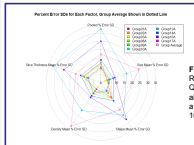
- Relative bias is within a tolerance of 15%.
- Scanner equivalence is found only for the larger synthetic lesions (10 mm and 20 mm). This finding confirms the lesion sizing guidance (10 mm and up) in QIBA CT imaging profile.
- Equivalence of the protocols supports the imaging protocol as used by ACRIN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

Recent Progress

QIBA 3A Group

Inter-algorithm Performance Investigation Study using FDA phantom Data

Aim: estimate absolute volumes using **CT phantom data.** Explicitly indicate descriptive statistics: **bias, variance.**



Study Results :

Figure 1 Percent Error for all Participants (without the Reference markups): Using only the nodules that is met QIBA CT profile, the standard deviation from pooled data for all 10 participants are shown by the dotted pink polygon with a pink dotted polygon. The pooled standard deviation of each 10 participant is shown by the different colors with a polygon.

Conclusion: The results support QIBA performance claims that the process of acquiring volume measurements according to the QIBA Profile should produce quantitative results where the upper endpoint of an approximate 68 % uncertainty interval for the systematic deviation between an estimate and the true value (i.e., +/- 1-sigma uncertainty level) with less than 15 %. Results also address the primary hypothesis that quantitative performance claims for tumor volume may be met with a variety of heterogeneous measurement algorithms ranging from semi- to fully automated methods.

QIBA 3B Group

Validation of VCT as a better imaging biomarker for predicting patient survival

Aim: Explore variability in measuring change in total tumor volume (uni and bi as well) and correlate tumor early responses assessed by the new volumetric technique and RECIST with patient survival in metastatic colorectal cancer (mCRC)

Study Results :

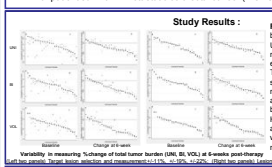


Figure 1. Waterfall plots of the baseline total tumor burden (TTB) in UNI (A), BI (B) and VOL (C) and their relative changes at 6-week (D-F) for each of the 29 individual patients. The solid circle, hollow circle and + sign represent radiologists 1, 2 and 3, respectively. Target lesion selection added considerable variability to baseline TTB measurements. However, relative changes in TTB among the three radiologists were within narrower ranges for both measurement models.

Variability in measuring %change of total tumor burden (UNI, BI, VOL) at 6-weeks post-therapy

(Left two panels: Target lesion selection and measurement) (-11%, +/-19%, +/-22%). (Right two panels: Lesion measurement only: +/-8%, +/-14% and +/-12% (Zhao)

Conclusion: The study found that variability in measuring early change of total tumor burden was low: 22% when considering both target lesion selection and measurement and 12% when only considering lesion measurement. The relatively low variability in change of mCRC measurements suggests that conventional response criteria could be modified to allow more accurate and sensitive CT assessment of anti-cancer therapy efficacy.

Zhao B, Lee S, Lee HJ, Tan Y, Qi J, Persigehl T, Motley PD and Schwartz LH. Variability in assessing treatment response: metastatic colorectal cancer as a paradigm. Clin Cancer Res. Published OnlineFirst on April 29, 2014; DOI: 10.1158/1078-0432.

Ongoing Efforts

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

QIBA 3A Group

We evaluated variability of scalar volume measurements, in terms of repeatability and reproducibility. We also compared segmentation boundaries relative to reference standard segmentations. An *important outcome of this work is the set of metrics* used to define performance for clinical CT data, needed in order to use volume change as a biomarker. *These metrics will form a basis for future determination of compliance with the QIBA Profile.*

The precision of the tumor volume measurements, characterized by both repeatability (within-lesion) and reproducibility (between-algorithm variability) was estimated. Relative magnitudes of various sources of variability were estimated using a linear mixed effects model. We also compared segmentation boundaries relative to reference standard segmentations to provide a basis on which to optimize algorithm performance.

The repeatability coefficient ranged from 6% (best) to 150% (least performing), corresponding to within-subject coefficients of variation of 2.1% to 54%. The reproducibility coefficient was 37%. Variability of smaller tumor volumes was lower without human editing, although larger tumors benefitted by editing the results. One-fifth to one-half of the total variability comes from sources independent of the algorithms. Over-segmentation occurred more often than under-segmentation. *Eight of the twelve participating algorithms performed at a level sufficient for QIBA compliance as judged on this data set. Based on these results, change in tumor volume can be measured with confidence to within +9% using any of the eight compliant algorithms.*

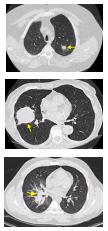


Figure 1: Examples of tumors used

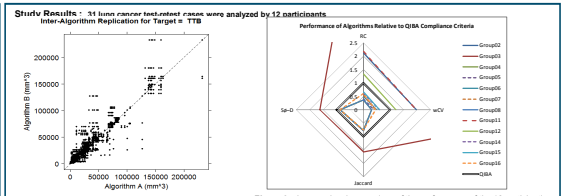


Figure 2 Results of Inter-algorithm analysis. Volumes for all pairs of measurements (same tumor, different algorithm) shown in the original scale. Bland-Altman analysis concluded Upper Agreement Limit=0.39 log(mm^3), and Lower Agreement Limit=-0.35 log(mm^3). These limits correspond to a range of 45% to 245%.

Figure 3 shows a visual comparison of the performance of the 12 participating. The plot captures all of the metrics evaluated in this study within a single plot and facilitates inter-comparisons among the algorithms. Best performance is closest to the origin of the plot, and below the "QIBA" line, suggesting that those groups plotted inside the line could be considered "compliant" whereas those outside the line could be considered "non-compliant". Missing points represent groups which did not submit segmentation objects.

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

Background:

•The QIBA CT Volumetry Biomarker Committee has completed its effort to define a process for optimal quantification of **advanced** lung cancer masses.

•The current task addresses analogous issues in the use of quantitative imaging to identify significant lung nodules in the CT-based detection of **early** lung cancer during the screening process.

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

Rationale:

•CT provides an effective means of detecting and monitoring lung nodules; CT screening can lead to a reduction in mortality in individuals at high risk for lung cancer.

•Size quantification on serial imaging is helpful to evaluate whether a nodule is benign or malignant.

•Quantification of nodule volume (volumetry) could solve some of the limitations of current diameter-based measurement metrics.

The Challenge:

• CT screening presents a challenge for the development of an optimized protocol in the need to balance the risks and harms in an asymptomatic population: lowest dose possible vs. reduced image quality that would limit the ability to detect small nodules.

• Any protocol will represent a compromise between these competing needs in the screening setting.

• Volume measurement error rapidly increases and repeatability rapidly decreases the further nodule size decreases below 1 cm.

The QIBA Profile:

•will make claims about the confidence with which changes in lung nodule volume can be measured under a set of defined image acquisition, processing, and analysis conditions.

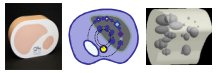
•will provide specifications designed to allow the achievement of targeted levels of clinical performance in identified settings within the screening context.

•will be intended for all stakeholders invested in lung cancer screening, including designers of CT image-acquisition protocols, those involved in the procurement of new CT equipment, software and device manufacturers, pharmaceutical companies, clinicians engaged in the screening process and in clinical trials, those who obtain quantitative measurements on CT images, and those who make decisions based on quantitative image measurements.

Funded Projects in Early Stages

Phantom Study of CT Volumetry for Hepatic Metastasis

- Aim: Assess performance of lesion sizing tools in estimating volume of synthetic hyper- & hypo-dense liver lesions
- Methods
  - Custom QRM-Liver® Phantom
    - Arterial phase & portal-venous phase inserts



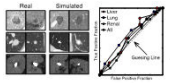
  - Synthetic Nodules
    - Shapes: spherical, lobulated, elliptical
    - Size: 6-40 mm

- Study Design
  - CT scans
    - Range of CT vendor platforms, exposures, slice thickness & recon parameters
  - Volume estimation algorithms
    - Columbia segmentation-based algorithm
      - Volume & lesion boundaries
    - FDA matched filter based algorithm
      - Volume estimates only
- Analysis
  - Linearity & bias
  - Repeatability & reproducibility
  - ANOVA identifying important image/lesion factors affecting volume estimation

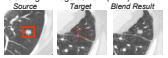
Reference Image Dataset for CT Volumetry with Known Ground Truth

- Motivation
  - Physical phantoms are useful, but often lack realism and variability
  - Clinical cases often lack ground truth
- Aims
  - Compare methods to insert lung lesions of known shape and volume into clinical CT images
    - Year 1: Establishing technique and influencing parameters based on physical phantoms
    - Year 2: Develop a dataset of clinical CT scans with synthetic lesions for quantitative verification

Simulated 3D liver, lung, and kidney stone lesions indistinguishable from real lesions based on expert ROC studies.



- Methods
  - Insertion techniques
    - A) Lesion mask → Projection → Recons. (Duke)
    - B) Lesion blending in image domain (FDA)



  - Dependency
    - Assess effect of acquisition parameters, segmentation, nodules shape and size, and vessel attachment
  - Calibration and Validation
    - Compare inserted lesion volume estimates to truth
    - Adjust Technique B for maximum concordance with Technique A and the gold standard
- Deliverables
  - Year 1: Lesion insertion software, phantom database
  - Year 2: Hybrid reference dataset of 100 clinical CT

Profile Development

- Profile structure:
  - Section 1: Background
  - Section 2: Claim – What performance can be achieved if profile is followed
  - Section 3: Procedures – How to achieve the performance in the claim
  - Section 4: Compliance – How to determine whether the procedures have been followed
- Current version of profile has gone through public review and been implemented in limited settings. Find the latest version on the QIBA wiki :
  http://qibawiki.rsna.org/index.php?title=Volumetric-CT_tech_ctte
- Ongoing revision
  - Restatement of claim: balance simplicity, clinical utility, and statistical rigor
  - Re-examination of procedures in light of claim
  - Preparation of checklists to establish compliance
    - Activities: Patient preparation, scan acquisition and reconstruction, image quality checks, image segmentation and analysis
    - Actors: Scanners, technologists, radiologists
- Next step: "field test" to move towards technical validation
  - Show that profile can be followed in practice

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.