

HHSN268201500021C

Quantitative Imaging Biomarkers Alliance (QIBA)

FINAL PROGRESS REPORT: AS OF MARCH 2018

This final progress report is stated in terms given in the accepted Work Plan, and is organized in the following subsections:

- A. Review of activities responsive to each objective.
- B. Combined list of groundwork projects associated with the objectives approved by the Steering Committee.
- C. Additional descriptions of general committee progress.

A. REVIEW OF ACTIVITIES RESPONSIVE TO EACH OBJECTIVE.

An update on Objectives 1-7 is given below. Note that, in general, the data from each of these efforts have been or will be uploaded to the Quantitative Imaging Data Warehouse (QIDW) identified in the methodology for Objective 6.

OBJECTIVE 1. CREATE AND DISSEMINATE NEW PROTOCOLS AND QIB PROFILES EACH YEAR THAT ADDRESS DISEASES OF SIGNIFICANT BURDEN TO THE US POPULATION.

Selected specifics:

CT Volumetry Biomarker Committee:

The CT Volumetry Profile for volume change in advanced disease has achieved the Consensus Stage. Feasibility tests have been conducted at a small group of clinical centers to assess its use in the field, and the feedback from these tests has been used to update the Profile.

Groundwork projects focused on volumetry of liver masses have been completed for the purpose of extending the first version of the CT Volumetry Profile to include quantitative assessment of liver masses.

Small Lung Nodule Biomarker Committee:

The committee has made numerous revisions to all sections of the Profile. The key elements of the current version of the Profile have been recently published (*Rydzak CE, Armato SG, Avila RS, Mulshine JL, Yankelevitz DF, Gierada DS, Quality Assurance and Quantitative Imaging Biomarkers in Low Dose CT Lung Cancer Screening, British Journal of Radiology, In Press*) to complement our earlier Profile-related publication in the *Journal of the American College of Radiology*. Publishing in a European imaging journal allows dissemination of this QIBA approach internationally and potentially gives greater incentive to international imaging vendors to integrate the QIBA quality imaging process. Two pilot efforts, one in the United States and one in international sites, are using a novel phantom designed for small lung nodule volume measurement to be evaluated with a proposed conformance process that utilizes cloud-based automated analytics. Analysis of this experience will allow for further refinement of the Profile.

CT Lung Density Biomarker Committee:

A Profile for the measurement of lung density based on quantitative CT measurements is in development and is expected to be released for review/approval by the CT Coordinating Committee in 2Q2018. Multiple factors are under consideration in the Profile draft, including improved harmonization of density measures across CT scanner makes and models, CT dose reduction, and improved lung inflation volume correction (2015-16 Round-5 Project L) to improve repeatability of density measures.

A Round-6 groundwork project (2016-18 (NCE) Project U) has been completed for feasibility testing of the CT dose reduction and harmonization approaches developed in previous cycles. The proposed methods for improving repeatability of density measures have been implemented and tested in repeatability studies as part of the COPD Gene Study, an ongoing, NIH-funded, multi-center longitudinal clinical research study of COPD disease severity and progression. This work directly informs new versions of the Profile with respect to

QIBA Final Report - March 2018

automatic exposure control (AEC) and iterative reconstruction (IR) as results of repeatability studies become available.

fMRI Biomarker Committee:

The QIBA BOLD fMRI Profile v1.0 titled *Mapping of Sensorimotor Brain Regions using Blood Oxygenation Level Dependent (BOLD) Functional MRI as a Pretreatment Assessment Tool* was released for public comment. RSNA staff members sent notice to relevant organizations, as well as QIBA membership, welcoming comment on the Profile. For these projects, fMRI Digital Reference Objects (DROs) were used to investigate the potential effects of head motion on the Profile claims and to help characterize data quality control metrics (e.g., minimum acceptable head motion) that form the basis for subject image data QC and Profile conformance. These DROs are available on the QIBA Quantitative Imaging Data Warehouse (QIDW). Definition and standardization of means to measure activation “center of mass” (COM), the Profile measure, have been identified as a potential issue since COM may not be readily measured by sites using their preferred fMRI analysis package(s). Even if software packages claim to report COM, the algorithm behind COM metrics may vary. Clear definition and technical steps to measure COM are refined in Profile Activities (Section 3) and Conformance (Section 4) of the Profile. While the Sensorimotor Profile was under public review, the fMRI BC began work on language function mapping based on a Round-5 (2015-16) groundwork project, currently being drafted into a Language Mapping Profile. The Language Mapping Profile development was supported by Round-5 (2015-16) Projects D1 and D2, completed in May 2017. Profile conformance remains an issue to be addressed; real-world feasibility testing is needed to determine if imaging sites would invest the extra time/effort outlined in the Sensorimotor Profile, with greater challenges anticipated for the Language Profile 2.0. The fMRI BC committee is drafting a Sensorimotor Profile Checklist for associated conformance procedures.

Perfusion/Diffusion/Flow (PDF)-MRI Biomarker Committee:

The Dynamic Contrast Enhanced (DCE)-MRI Task Force has ported elements of DCE Profile v1.0 into the new Profile template for revision into DCE-MRI Profile v2.0, which will extend the Profile to include 3.0T systems and parallel imaging techniques commonly used in DCE studies. Most Task Force efforts have been focused on an organ-specific, systematic literature review process and groundwork projects. As with other biomarkers, the DCE TF has noted that despite a large volume of publications, there are relatively few good quality test-retest studies essential to development of the Profile claim(s). Candidate DCE organ sites reviewed for development of claims include: brain, head and neck, prostate, and breast. Claim statements for a “true change in K_{trans}” were developed for brain (GBM) and prostate. The Profile Activities and the Assessment Procedures (sections 3 and 4, respectively) are being redrafted for consistency with other QIBA Profiles. A PDF Biomarker Committee-approved version of DCE Profile v2.0 is anticipated 2Q2018.

Extensive rewrite of the Diffusion-Weighted Magnetic Resonance Imaging (DWI) Profile v1.0 was completed by the DWI Task Force in 4Q2016, and the claims were expanded to include prostate (along with brain and liver organ sites). The DWI Profile was released for the public comment in April 2017 and the public comment period closed on August 25, 2017. Currently, the DWI TF is addressing comments received for their incorporation by 2Q2018. A Checklist for the DWI Profile is also being developed.

Two PDF-MRI Biomarker Committee Task Force efforts focused on completion of ground work projects as well as early-stage of Profile development: 1) The Diffusion Tensor Imaging (DTI) Task Force, co-chaired by Drs. Provenzale (clinical lead) and Schneider (technical lead), is distilling a substantial volume of DTI publications to identify suitable biomarkers for the main target organ, which is the brain. Standard fractional anisotropy (FA) and other scalar metrics, along with more advanced metrics derived from higher directional and multi-shell “DTI+” are being considered. 2) The Profile development for Dynamic Susceptibility Contrast (DSC) is led by Drs. Erickson and Wu. Elements of the DSC Profile have been drafted into the current QIBA Profile Template. Technical assessment procedures are pending phantom development (Round-5 ground work project H) and multi-site/vendor measurements to inform the DSC Profile as well as DSC DRO development (Round-6 ground work project Y). A DSC Profile version ready for PDF Biomarker Committee and MR Coordinating Committee review is expected in 3Q2018.

Formal collaboration between the European Imaging Biomarkers Alliance (EIBALL) and QIBA on development on an Arterial Spin Label (ASL) neuro perfusion Profile began 1Q2017. Essentially all ASL committee activities

QIBA Final Report - March 2018

are centered in Europe (EU dominated membership) with administrative support provided by the European Institute for Biomedical Imaging Research (EIBIR), although periodic updates are presented to the QIBA PDF-MRI committee.

MR Elastography (MRE) Biomarker Committee:

The MRE committee completed its Profile v1.0 draft titled *Magnetic Resonance Elastography of the Liver* based on an extensive meta-analysis of the test-retest repeatability of hepatic MR elastography literature. The Profile was released for public comment in August 2017, and the public input phase closed on September 15, 2017. The biomarker committee is now reviewing and addressing the public comments.

Proton Density Fat Fraction (PDFF) Biomarker Committee:

This committee is actively developing a draft Profile v1.0 for quantitative proton density fat fraction measurements in liver. Since substantive groundwork has already been performed and results previously published, Profile development is progressing well. A robust literature search was completed, which informed claim design, and the Profile is in draft phase exploiting significant format commonality with the recently completed MRE Profile. Further, a meta-analysis that is being used to support cross-sectional and longitudinal claims was completed and published in *Radiology*. A first draft of the Profile is expected early 2018.

Musculoskeletal (MSK) Biomarker Committee:

This recently-formed committee, co-chaired by Drs. Thomas Link and Xiaojuan Li, includes international membership and is directed to develop Quantitative Compositional MRI relevant to cartilage, namely T1rho/T2 relaxation times as non-invasive imaging biomarker of joint disease. Many of the current MSK BC efforts are focused on increasing the visibility of QIBA's MSK quantitative efforts to better engage non-imaging clinicians, e.g., rheumatologists, osteoarthritis societies and foundations. A meta-analysis reproducibility paper on cartilage compositional biomarkers will be published in *Osteoarthritis and Cartilage*; Drs. Li and Link submitted an editorial concerning this paper. Discussions are underway with High Precision Devices (HPD), Inc. and NIST to develop a physical phantom for quantitative T2 and T1rho. Several committee members are looking to leverage funding opportunities (R01 and P50 awards) to advance MSK BC objectives.

FDG-PET/CT Biomarker Committee:

The *FDG-PET/CT as an Imaging Biomarker Measuring Response to Cancer Therapy* Profile advanced to the Technically Confirmed stage November 30, 2016. This is version 1.13, dated November 18, 2016, available at <http://qibawiki.rsna.org/index.php/Profiles>. To be a Technically Confirmed Profile, the Profile details have been implemented in more than one facility using equipment from more than one vendor and each individual actor (hardware, software and person) successfully met the specifications. This was verified through two rounds of feasibility-testing across multiple sites and the FDG-PET/CT Biomarker Committee asserts that deployments will find the Profile requirements practical/feasible to understand and execute. Planning is now underway for a clinical trial to move the Profile to the Clinically Confirmed stage. An application to the ACR Foundation for funding has been submitted.

PET-Amyloid Biomarker Committee:

The QIBA Profile *18F-labeled PET Tracers Targeting Amyloid as an Imaging Biomarker* completed the public comment phase, and comments received through the September 2017 end date have now been addressed. The Profile is close to being voted as a Consensus Stage version and Biomarker Committee members are planning feasibility/conformance testing for the Profile and contacting possible testing sites. A second-generation Digital Reference Object (DRO) has been designed and was tested for the quantitative assessment of amyloid tracers. Based on these results, a second version is being developed. The committee has agreed on their Profile claim based on current literature citations, as well as data from Round-5 (2015-16) Project E.

SPECT Biomarker Committee:

The first SPECT Profile (*Quantifying Dopamine Transporters with 123-Ioflupane in Neurodegenerative Disorders*) addresses Parkinson's disease (PD) and has advanced to the Publicly Reviewed stage as of June 2017. The public comment phase continued until the end of February 2017, at which point the SPECT

QIBA Final Report - March 2018

Biomarker Committee revised the Profile to respond to the public comments, which is now complete. It is worth noting that (1) this first draft of the Profile advanced to the Version for Public Comment stage in less than one year, and (2) substantive public comments have been received already. International enthusiasm for participation has been particularly strong from Japan and several European Union states.

The SPECT Biomarker Committee has also started to draft a second Profile for 99m-Tc SPECT in Immunology/Oncology.

US SWS Biomarker Committee:

A complete draft of the Profile *Ultrasound Measurement of Shear Wave Speed (SWS) for Estimation of Liver Fibrosis* was circulated to SWS Biomarker Committee members on February 2, 2018 for review before the first public circulation, estimated May 1. Checklists in the current draft form the basis for the self-attestation of the conformance program newly established for the Profile. These and other recent additions to the Profile have been informed by Round-5 Projects F1 and F2 and Round-6 (2016-17) Project BB. Mitigation strategies to reduce SWS measuring system variance in clinical procedures have been developed based on the results of the Phase II studies in phantoms and Round-6 Project BB. These strategies and new methods of extrapolating US shear wave speed imaging results to the lower frequencies employed in MRE and the established Fibroscan system will appear in the next version of the Profile.

An example claim follows:

CLAIM 1 (technical performance): A shear wave speed measurement has a within-subject coefficient of variation (wCV) depending on the measured SWS and depth of acquisition according to the following table:

Measured SWS	Coefficient of Variation (X)	
	Depth=4.5	Depth=7.0
SWS<1.2	5%	8%
1.2 < SWS < 2.2	4%	5%
SWS>2.2	10%	12%

This claim is based on a phantom study which demonstrated the following results.

Precision (repeatability) Estimates from Phantom Data

Mean SWS	wSD (m/sec) [N]		wCV	
	Depth=4.5	Depth=7	Depth=4.5	Depth=7
<1.2	0.040 [134]	0.050 [141]	4.2%	6.6%
1.2 – 2.2	0.063 [122]	0.087 [136]	3.0%	4.2%
>2.2	0.132 [13]	0.189 [9]	5.8%	8.5%

US Volume Blood Flow (VBF) Biomarker Committee:

Work on the groundwork for the profile and systems to follow the Profile continues. Four manufacturers are actively improving or modifying their platforms to provide volume flow *via* the QIBA algorithm with the goal of its introduction into the clinic. Two of the three systems tested with full analysis exhibit a mean +2% bias and one had a ~ -35% bias. The latter is due to a simple, correctable error in data output from the system. The average coefficient of variation is 4.6%. Six platforms are providing, or in preparation for providing, data for the measurements.

CEUS Biomarker Committee:

Since the Biomarker Committee is relatively new, it has not started outlining the Profile. It has had no subcontract for groundwork projects. However, the committee is active and national and international volunteer efforts by industry, medical, medical support and biomedical researchers have been strong.

OBJECTIVE 2. PERFORM FIELD TESTS AND REVISE EXISTING QIB PROFILES AS NEEDED.

Selected specifics:

CT Volumetry Biomarker Committee:

The advanced disease Profile has been tested at three sites. Feedback has been collected and the feedback has been used to revise the conformance procedures.

fMRI Biomarker Committee:

Committee discussion has centered upon planning staged, multi-site, human testing of the Profile (after receiving and incorporating public comments). The first stage will involve multiple sites within the Biomarker Committee's membership implementing Profile v1.0 and acquiring human subject motor mapping data. Testing will focus on identifying implementation issues, subject QC methods, and agreement with claims in cases where multiple measurements are made in the same subject. The DROs will be used to find out if the performance plan is feasible. One scenario is to create and share a DRO of 20-30 cases known to surpass QC thresholds, and provide a step-by-step assessment procedure. This range of DROs is designed get the center of mass (COM) within Profile specifications. The next stage will likely involve extending such testing to additional sites that perform clinical fMRI motor mapping, where site-generated datasets would undergo the same assessment procedure. Checklist and assessment procedures were drafted wherein the Profile user needs to: 1) Be able to acquire data from the scanner that matches the QA measurement data within QIBA Profile specifications; 2) Demonstrate that COM calculation is within QIBA Profile specifications (can be assessed using DROs rather than acquired empirical data); and 3) Assess bias and repeatability using the data provided.

PDF-MRI Biomarker Committee:

Overseen by the PDF-MRI Biomarker Committee, the DCE-MRI Task Force is awaiting analysis of the *DCE-MRI Quantification* Profile (v1.0) field test (design supported by a Round-2 groundwork project), which was implemented in the ACRIN 6701 prostate cancer patient test/retest study. This study also included site certification by scanning physical DWI and DCE (T1) phantoms. The ACRIN 6701 study was specifically designed to collect multi-site/multi-vendor DCE (and DWI) data, and will thus directly inform future QIBA Profiles involving MR for prostate imaging. Accrual for this study has been completed, and preliminary results are anticipated in 2Q2018. As the DCE-MRI Profile v2.0 is recast into the new QIBA template, material related to 3T systems and parallel imaging techniques is being incorporated. Known B1 non-uniformity issues at high-field (3T) are being addressed in the 2016-17 Round-6 Project Z groundwork project, which involved building and testing 4 physical phantoms suitable for B1-and T1-mapping. Initial round of scanning head- and torso-sized B1-uniformity phantoms was completed on multiple vendor platforms (2017-2018) to directly inform the DCE Profile v2.0. Second round (repeatability) phantom scanning is ongoing.

Beyond phantoms used in ACRIN 6701, a multi-site/multi-vendor study of a polyvinylpyrrolidone (PVP) DWI phantom (supported by a Round-3 groundwork project) was performed with data acquisition completed mid-2016. In parallel, a commercial version of the QIBA DWI phantom was used for QC/site qualification in the "Track Traumatic Brain Injury" multi-center trial. Acquisition and analysis using a common DWI QA/QC software package (supported by 2013-14 Round-3 groundwork contract) directly informed technical Performance Assessment procedures and specifications contained in the DWI Profile.

The DTI Task Force has completed accumulating multi-site/platform DTI data using a novel phantom (supported outside of QIBA) based on "taxon" inserts to provide ground truth in anisotropy metrics. These physical phantom field tests constitute a base technical DTI performance assessment to inform conformance

QIBA Final Report - March 2018

procedures for the DTI Profile. Fractional Anisotropy (FA) repeatability/reproducibility data will be presented later in this report.

FDG-PET/CT Biomarker Committee:

The FDG-PET/CT Profile has completed two rounds of field-testing to evaluate its feasibility and practicality. The first round was at four academic imaging sites, while the second round was at 12 regional clinical sites across the USA. User-suggested changes to the Profile were made, as well as a revision of the checklists of conformance procedure steps. An updated version (Technically Confirmed Stage) of the FDG-PET/CT Profile was completed 4Q2016.

US SWS Biomarker Committee:

Procedural checklists and the rest of the draft profile have been utilized in obtaining human data for the Round-6 technical confirmation study at MGH and soon at the Washington DC VAMC. If the revisions based on full data analysis are not too extensive, the committee will attempt to not only reach the Technically Confirmed stage but also the Claim Confirmed stage of Profile development.

US Volume Blood Flow (VBF) Biomarker Committee:

Field tests on patients undergoing carotid MRI and volume flow ultrasound with catheter reference measurements are planned without support from this contract. If the results are as promising as expected, the committee will attempt to not only to reach the Technically Confirmed stage but also the Claim Confirmed stage of Profile development.

OBJECTIVE 3. PERFORM INDIVIDUAL GROUND WORK DATA COLLECTION AND ANALYSIS PROJECTS TO FILL GAPS IDENTIFIED DURING WORK DEVELOPING QIBA PROFILES COVERING THE FOUR MAJOR IMAGING MODALITIES, CT, MRI, RADIONUCLIDE, AND ULTRASOUND.

Selected specifics:

The Round-6 (2016-17) groundwork projects have been completed.

CT Volumetry Biomarker Committee:

Methodology and Reference Image Set for Lesion Characterization in terms of Texture and Morphology, PI: Ehsan Samei, PhD - Duke University (CT Volumetry Biomarker Committee) - Round-5 (2015-16) Project C

Previous work validated the basic methods for creating realistic “hybrid” CT images, in which computational lesion models with pre-defined characteristics (e.g., size, shape, and contrast) were virtually inserted into deidentified patient CT data. This was followed by establishing a level of exchangeability between real and simulated lesions for volume estimation tasks through a QIBA Grand Challenge (supplemental material for the QIBA Grand Challenge is attached below). The purpose of this project was to provide a first-order investigation into the fabrication of lesion internal heterogeneity. These databases can serve as surrogates to actual patient data, albeit with lesion size, shape, and texture classification known *a priori*. To this end, clinical data from the Lung Image Database Consortium (LIDC) was utilized to benchmark texture feature calculations. These chest cases contained confirmed malignant lesions, where 21 Haralick texture features were extracted and calculated. Based on this real spatial architecture, corresponding computational models were created to statistically model the features calculated from patient images using the 3D cluster-lumpy background algorithm (3D-CLB). 3D-CLB texture features were then calculated and compared to real texture using mahalanobis distance where features were objectively similar to within 13%. This indicated successful emulation of the texture properties of the real lesion in the texture synthesis. To determine the impact of imaging system parameters on texture feature measurement, a series of simulations based on a range of modulation transfer function (MTF), noise power spectrum (NPS), voxels size and ROI size conditions were simulated. Twenty-one Haralick texture features were calculated and compared between the ground truth phantoms and their corresponding vendor-specific blur and noise simulations using percent difference (PD).

QIBA Final Report - March 2018

Noise level impacted texture feature calculation strongly by as much as 500% across 10 out of 21 features. Features such as contrast and dissimilarity were highly different for simulated vs. ground truth. This effect varied among image systems. The two CT systems generally yielded different PDs for a given feature at the same voxel size and CTDI level.

Additionally, a framework to assess lesion morphology deformation was assessed. Regional Hausdorff Distance (RHD) was computed from two 3D volumes: computer-aided design (CAD) models of lesions and mesh models of CT imaged lesions. First, 24 physical lesions of 4 shapes (spherical, elliptical, lobular, and spiculated) and three sizes (8, 9, 10 mm) were inserted into an anthropomorphic Kyoto thorax phantom. The phantom was scanned with the Siemens Flash scanner across two CTDI_{vol} levels (1.45 and 22 mGy) and two reconstruction kernels (filtered-backprojection: FBP and iterative: IR). To model the impact of system blur on lesion shape, CAD models of physical lesions were virtually inserted using three previously-validated virtual insertion techniques from a Round 4 project. Both physical and virtual lesions were segmented from the image. The Hausdorff Distance (HD) was then calculated from each pair of physical and virtual lesions. The minimum HD value represented the closest matching pair. Finally, the 3D RHD map and the distribution of RHD were computed for the matched pair. The data were fit to a non-linear regression model, where the figures of merit were the mean, standard deviation and coefficient of variation of RHD, denoted $Mean_{RHD}$, STD_{RHD} and CV_{RHD} , respectively. When considering CV_{RHD} for insertion techniques, there were minimal differences (6%, 5%, 5%) between HD measurements for baseline conditions (8 mm spherical, CT-imaged lesion, at low dose, and reconstructed with FBP) and Techniques A, B, and C, respectively. Overall, these results indicate that the expected morphological changes due to the CT system were modeled similarly by the virtual insertion techniques, and can be closely modeled by incorporating the appropriate system MTF.

AIM 1 RESULTS AND DISCUSSIONS

When the scale indicates that PD is 0, simulated conditions (noise and blur) did not influence the feature of interest. Such PD = 0 results were more prominent in cases where small ROIs were used. We find that features such as contrast and dissimilarity are highly different for simulated vs. ground truth and are influenced by dose levels, and ROIs. This effect varies among image systems. On the other hand, features such as Energy, Sum Average, Texture Entropy, Long Run Emphasis, Low Gray-Level Emphasis, and Gray-Level Variance seem to be unaffected by noise, and blur. Noise level impacts Haralick texture feature calculation strongly by as much as (500%) across 10 out of 21 features.

The PD for each feature using the same voxel size for a single CT system remains mostly similar across all ROIs, however, small differences should not be ignored. Variations in feature intensity suggest that CTDI (noise level) impacts texture feature calculation. Also, the two CT systems generally yield different PDs for a given feature at the same voxel size and CTDI level. Further, larger voxel sizes seem to suppress inter-dose feature variation (inter-dose variability decreases as voxel size increases). These pilot results pave the way for a broad study to enable generalization for statistical evaluation of system impact on texture feature calculations. As a follow up, the previously mentioned indications among others will be tested in the form of a comprehensive study on the measurability of texture features.

AIM 2 RESULTS AND DISCUSSIONS

The estimated effects show higher values of $Mean_{RHD}$ for larger size lesions, with significantly higher results for 9 mm (p-value = 0.03) and marginally so for 10 mm (p-value = 0.08). The analysis also shows significantly higher $Mean_{RHD}$ values for elliptical and spiculated shapes relative to spherical. $Mean_{RHD}$ was significantly lower for Technique C (p-value = 0.002). For the two reconstruction algorithms and doses used, there was no significant effect of reconstruction method or dose on the lesion morphology when the virtual lesions were compared to the physical lesions.

Analysis shows higher STD_{RHD} values for other shapes relative to spherical, especially for spiculated (p-value < 0.0001). This suggests a greater degree of difference between the idealized and the segmented morphology for other lesions compared to that for spherical lesions. A significantly lower STD_{RHD} for Technique C (p-value =

QIBA Final Report - March 2018

0.01) was observed. The effect of dose was relatively small but significant (p -value = 0.003) in the analysis of STD_{RHD} . In the analysis of CV_{RHD} , results show higher CV_{RHD} values for other shapes relative to spherical, especially spiculated lesions. Using the two reconstruction methods, there was no significant effect of reconstruction method on any of the measurement parameters.

Regional deviations from lesion true morphometry can be lost in a basic volume estimation. To address this concern, capturing 3D distortion can be most appropriate. Note that the $Mean_{RHD}$ indicate size differences, while STD_{RHD} indicate shape differences. Therefore, for quantifying changes in shape independent of size scaling, CV_{RHD} is the recommended figure of merit. When considering CV_{RHD} for insertion techniques, there were minimal statistically significant differences between the HD measurements and the baseline (8 mm spherical, CT-derived, low dose, FBP). Overall, these results indicate that the expected morphological changes due to the CT system were modeled similarly by the virtual insertion techniques, and can be closely modeled by incorporating the appropriate MTF. As expected, a greater degree of morphological changes was observed with spiculated lesions when compared to other lesion types.

AIM 3 AND 4 RESULTS AND DISCUSSIONS

A database composed of 100 patient thoracic images and 100 patient abdominal images was collected from Duke University Hospital. Using the Duke database of lesion masks and their cloned renditions, 80 lesion morphologies (ground glass, part-solid, and solid with varying shapes and contrast and with sizes ranging within 4-20 mm) were virtually inserted into each patient lung image set. All lesions contain realistic internal heterogeneity. Similar insertions were applied to liver images. Patient images and lesion models were examined by experienced radiologists. This database is packaged to be uploaded to the QIDW.

A software package was created to insert lesions dynamically into clinical datasets. The software has the functionalities to provide lesion addition and database creation upon demand. While the static database of Aim 3 enable conformance with pre-defined performance thresholds, this software can offer the similar goal using databases that are known only statistically and thus any testing and training can be done in a more objective manner.

CT Lung Density Biomarker Committee:

Investigation of Methods of Volume Correction for Lung Density CT, PI: Sean Fain, PhD - University of Wisconsin-Madison (Lung Density Biomarker Committee) - Round-5 (2015-16) Project L

A reference library of CT lung density histograms has been created from test-retest scans in human subjects (from the COPDGene and NLST studies) to assess consequences of inconsistent breath-hold on CT density measures in the lungs. The performance of previously published statistical models for lung volume adjustment of CT density measures was tested using these histograms. In addition, several imaging phantoms, consisting of standardized reference material (lung equivalent density foams) ("QIBA-SRM" phantom) and a piston system with shredded foam to enable controlled experiments at varied volume and density as a means to independently validate the different lung volume adjustment approaches in the literature, were implemented. The project results indicated statistical models that enforced subject specific slope = 0 but non-zero intercept performed best with respect to reducing the repeatability coefficient leading to this recommendation for longitudinal studies.

CT Lung Density Biomarker: Translating Phantom Harmonization to Clinical Practice, PI: Stephen Humphries, PhD - National Jewish Hospital (Lung Density Biomarker Committee) - Round-6 (2016-18 (NCE)) Project U

This multi-site, multi-vendor assessment is ongoing and compares repeatability for specifications using conventional and low-dose protocols that conform to the Profile claims. This is an important test of the technical feasibility of the draft Profile procedures in an ongoing clinical research trial that will specifically address the open issues of the impact of AEC and IR on density measures and their repeatability. In addition, the proposed correction technique derived from phantom studies could potentially reduce or eliminate the variation due to scanner make and model in patient data, which would be a major advance in quantitative CT of the lung.

QIBA Final Report - March 2018

Conclusions and future work: Prior rounds of phantom analysis have demonstrated the technical feasibility of using NIST-certified SRM foams to harmonize CT measurements from different scanners. Translation to clinical practice faces additional challenges. We observed modest improvement in associations between qCT lung density measures when a phantom-based correction factor was used. Reduced dose scanning protocols present additional challenges for harmonization. Iterative reconstruction can improve concordance between qCT measures derived from reduced dose and fixed dose acquisitions. The Lung Density Biomarker Committee will consider these results in making recommendations on scanning protocols for clinical applications relevant to the Profile.

fMRI Biomarker Committee:

Quantitating Clinical fMRI Mapping of Language: Center, Spatial Extent, and Relative Strength of Active Areas, PI: James Voyvodic, PhD - Duke University (fMRI Biomarker Committee) - Round-5 (2015-16) Project D1 and D2

We identified and analyzed 775 fMRI scans of language function from 355 subjects (retrospective data regarding patients and healthy controls), each of whom underwent more than one language scan. Of these, 340 subjects had multiple scans within the same session and 15 subjects had language scans acquired in different sessions. Two different language tasks had been performed. All subjects had a "Sentences" task; 260 had multiple Sentences tasks whereas 95 subjects performed both a Sentences task and "Words" task. This allowed us to evaluate reproducibility of language mapping both within and across tasks. Affine registration transforms were generated to register brain images from each session to a standard MNI reference brain and to other scan sessions of the same subject. Quantitative quality assessment measures were generated for all scans. These included head motion indices, consistency indices for task performance, overall BOLD activation metrics (mean and peak amplitudes, spatial extent), regional activation statistics for multiple putative language ROIs, hemispheric lateralization indices, and subjective mapping assessments by multiple raters. AMPLE-normalized language fMRI maps were generated and resampled to the common MNI anatomical space so different language task maps could be overlaid and measured for reproducibility. Reproducibility metrics included 3-D location of activation peaks, spatial extent of activation, and hemispheric language lateralization indices. The final step will be to evaluate these reproducibility metrics as a function of the QA metrics in order to identify new Profile claims for reproducibility of language mapping and the data qualification necessary to achieve those claims.

PI: Jay Pillai, MD - Johns Hopkins University (fMRI Biomarker Committee) - Round-5 (2015-16) Project D2
Two commonly-used clinical language fMRI paradigms have been evaluated in a group of >50 patients over a course of four years to assess both reproducibility within a single scan session and effectiveness of hemispheric lateralization using threshold dependent and independent methods. In the reproducibility assessment, we have thus far evaluated holohemispheric laterality indices (LI) and plan to evaluate region-specific LIs as well as correlate the findings with QC metrics for the remainder of the project. In addition, Dr. Pillai's fMRI DRO data were uploaded to the QIDW as the final deliverable for this Round-5 project.

PDF-MRI Biomarker Committee:

DWI-DRO Development for ADC Analysis, PI: Dariya Malyarenko, PhD - University of Michigan (PDF-MRI Biomarker Committee / DWI Task Force) - Round-5 (2015-16) Project G

The goal of this project was to provide a DWI DRO containing "modeled ground truth" with realistic Rician noise conditions for evaluation of diverse software packages that purport to convert DWI into quantitative ADC. This DWI DRO was modeled after the DCE DRO (Round-1 groundwork project) and utilized diffusion DICOM attributes defined according to the standard (vendor-independent) DWI "macro"
(http://dicom.nema.org/medical/dicom/current/output/chtml/part03/sect_C.8.13.5.9.html).

All stated project deliverables were met and included: a) definition of a wide tissue-relevant ADC/SNR parameter space, b) adherence to DICOM-compliant diffusion attributes, c) DWI generation based on the standard (mono-exponential diffusion model, though the framework is flexible for DRO extension to other tissue models in the future, d) a trial of the DRO using a select set of DWI analysis software packages, and e) delivery of DRO datasets with software performance analysis documentation to RSNA-QIBA for its distribution

QIBA Final Report - March 2018

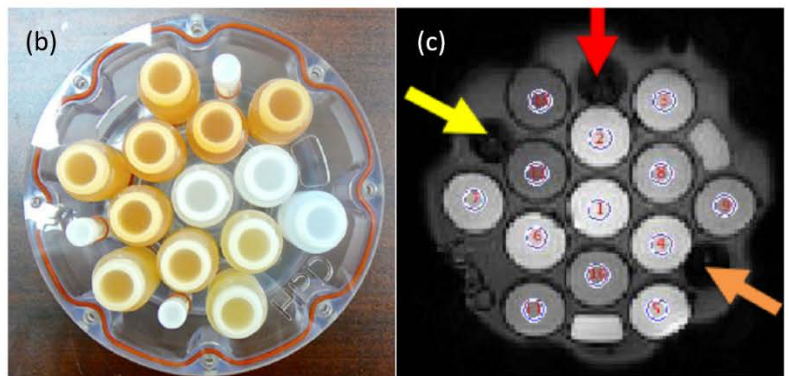
(on the Quantitative Imaging Data Warehouse, QIDW). As part of a collaboration with the NCI's Quantitative Imaging Network (QIN), QIBA DWI DRO development tools were modified by Dr. Malyarenko to extend to non-Gaussian diffusion to encompass IntraVoxel Incoherent Motion (IVIM), Stretched Exponential (SE), and Kurtosis Model (KM). Use and description of the non-Gaussian diffusion DRO can be found at: <https://www.researchgate.net/publication/322404473>. This non-Gaussian DWI DRO is being used to provide initial validation of QIN tools of diffusion kurtosis in prostate DWI.

Dynamic Susceptibility Contrast MRI Phantom, PI: Ona Wu, PhD - Harvard University / Massachusetts General Hospital (PDF-MRI Biomarker Committee/DSC-MRI Task Force) Round-5 (2015-18 (NCE)) Project H

The primary goal was to develop a prototype physical DSC phantom from which a gradient of susceptibility values can be measured. A secondary goal was to generate generic acquisition protocols by which one can assess the contrast-to-noise ratio of the susceptibility measurements, as well as stability across time and vendors. Finally, the third goal was to estimate reproducibility and feasibility of performing these measurements across multiple centers at multiple time points.

(a) Concentration of agarose and GdCl₃ for each vial

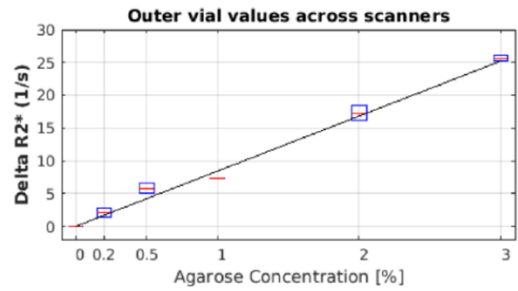
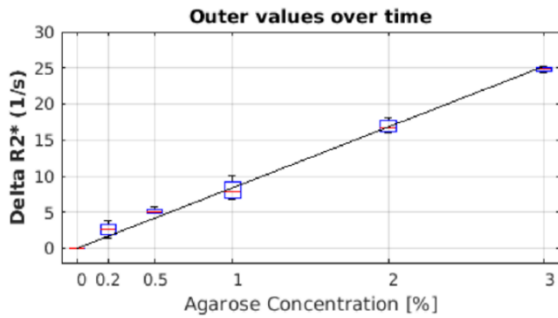
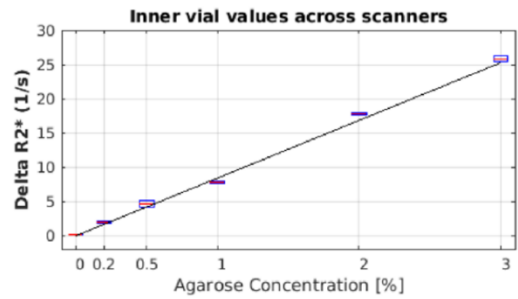
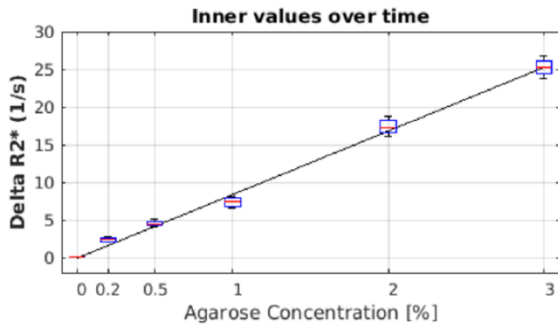
Vial	Sample (% agarose)
1, 2, 3	0.047 mM MnCl ₂
4, 5	0.01 mM GdCl ₃ + 0.2%
6, 7	0.01 mM GdCl ₃ + 0.5%
8, 9	0.01 mM GdCl ₃ + 1.0%
10, 11	0.01 mM GdCl ₃ + 2.0%
12, 13	0.01 mM GdCl ₃ + 3.0%



(a) DSC phantom composition (b) shell and 13 vial placement. Also shown is an axial MRI view (c) of the central slice of the DSC phantom. Reference tubes (arrows) filled with GdCl₃ and "HPD" label confirm phantom orientation.

Two phantom prototypes to estimate reproducibility across imaging sites were designed. A high-level generalized imaging protocol to be utilized with the phantoms was also developed. Candidate phantom components were tested for stability and suitable magnetic properties, such as susceptibility range and T₁ with the final composition shown in above. Software tools were developed for analyzing data based on prior NIBIB-supported (DWI) groundwork projects and used to evaluate repeatability shown below.

QIBA Final Report - March 2018



Boxplots of average $\Delta R2^*$ values for each vial measured 4 times over a period of 10 months. The range of values are similar to spatial variance shown in the figure above and can be due to variation in placement of ROIs and phantom positioning.

Boxplots of average $\Delta R2^*$ values for each vial in the phantom measured on two different scanners. The range of values are similar to spatial variance shown in the figure above and can be due to variation in placement of ROIs and phantom positioning.

A Web-based Tool for Creating DSC Digital Reference Objects, PI: Bradley Erickson, MD, PhD - Mayo Clinic, Rochester MN (PDF-MRI Biomarker Committee/DSC-MRI Task Force) Round-6 (2016-17) Project Y

The purpose of this project was to provide a web-based interface to several software packages that model DSC signal changes thereby creating images to simulate DSC perfusion data. The goal was to generate a 3D anthropomorphic head model to produce 4D data for each class of 'tissue' in the simulated object. The plan was to simulate normal white matter, normal gray matter, CSF, non-brain/muscle, air, vessels, and 4 'tumors' where the user may select the amount of leakage, relative blood volume, transit times, and dispersion effects. At this point, the Mayo computation module is running and the investigators have demonstrated scalability and management and are working on incorporation of other model modules. The software allows simulation of user-defined DSC scan protocols via pull-down menu of key parameters (see below) for DICOM DSC output where "truth" for CBV and tumor leakage are known.

BNIDRO

[show info](#)

Start service interactively

Command line parameters for **BNIDRO** . Typically these are strings or numbers.

B
Field Strength (Please choose between 1.5T and 3T)

Options ▾

FA
Flip angle (Please choose between 30, 60 or 90 degrees)

Options ▾

TR
Repetition Time (Please choose between 1000, 1500, 2000 msec)

Options ▾

echo
Option for Echo Time (Four options available, 1,2,3,4 corresponding to 20, 30,40 and 50 ms)

Options ▾

dosingscheme
Dosing scheme (Currently four dosing schemes are available 1, 2, 3, 4 and 5 corresponding to 0+1, $\frac{1}{4}+\frac{3}{4}$, $\frac{1}{2}+\frac{1}{2}$, $\frac{1}{4}+1$, 1+1)

Options ▾

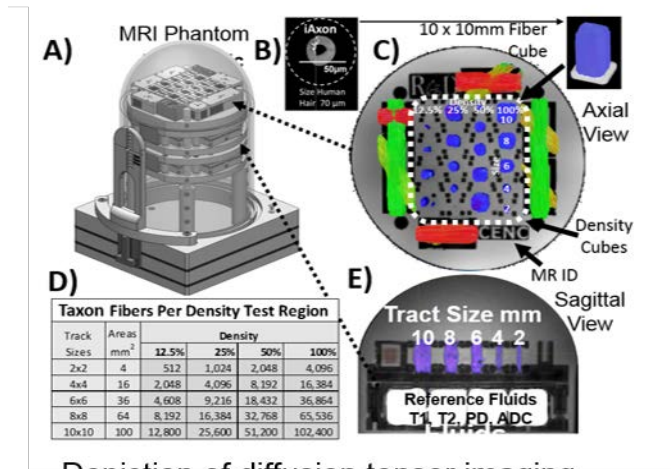
File name for files created by **BNIDRO** . These are individual files (not zip files) and should reflect any extensions that are important to the service.

output
Suggested -> output.zip

Measurements of Reproducibility of DTI Metrics on Clinical MR Scanners Using a DTI Phantom, PI: James Provenzale, MD - Duke University (PDF-MRI Biomarker Committee/DTI Task Force) Round-6 (2016-18 (NCE)) Project T

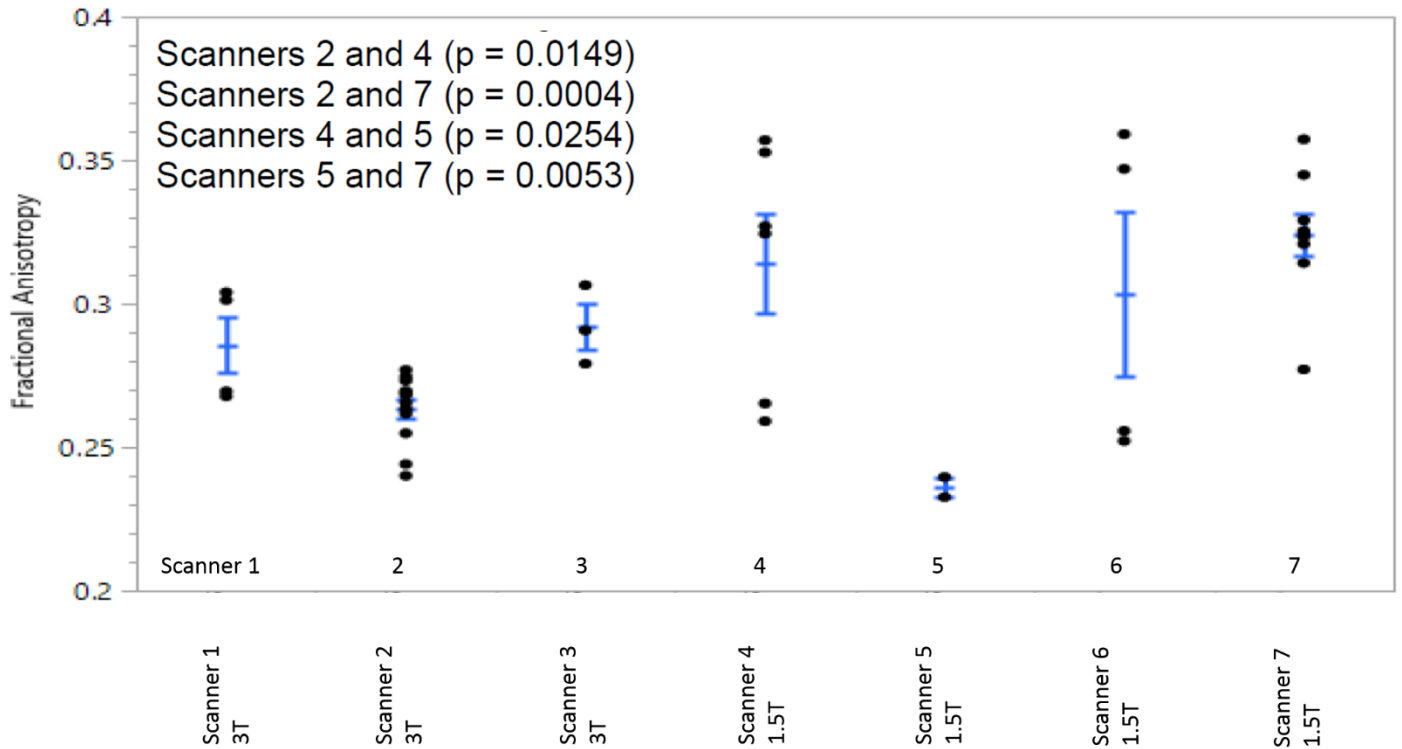
The goal of this proposal was to acquire the data needed to inform the QIBA DTI Profile using a novel phantom on representative multi-platform systems at a few medical centers. The phantom design (below) is stable over time and can provide “ground truth” measurements of DTI metrics, *i.e.*, mean diffusivity (MD), fractional anisotropy (FA), radial diffusivity (RD) and axial diffusivity (AD).

The project specifically addressed issues of acquisition protocols, vendor platforms, and field strength impact on production of FA values. The emphasis was on reproducibility of specific MR scanner models and specific imaging protocols. We performed 38 DTI scans on seven scanners (average: 5.3 scans, range 2-12) over a period of 12 months using a b value of 1,000 s/mm² and 12 encoding directions.



Depiction of diffusion tensor imaging phantom used in this project.

Data on 4 systems (figure below) indicate large systematic bias in FA for one system “C” (left) is reduced by corrective re-calibration (right).



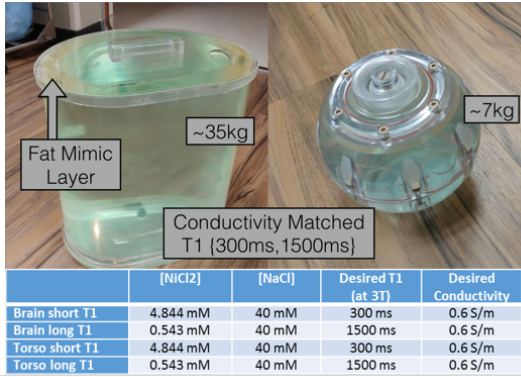
Substantial variance in DTI phantom FA values (above) was observed across scanner platforms. Degree of statistical differences in the mean FA for scanner pairs is shown in the inset.

Using this complex phantom as a reference standard for FA values, substantial inter- and intra-scanner variability in MR scanner performance was clearly demonstrated. In Part 1, the evaluation of combined data from all MR scanners and all imaging protocols, we found that great variability of DTI data is evident when MR scanners of different field strengths, different MR vendors and different numbers of diffusion gradient directions are used. These findings are expected; however, they are a reminder that compiling DTI data from multiple sources runs a high likelihood of producing confusing and misleading findings.

Our findings indicate that reproducibility of FA values is often relatively low, even on the same MR scanner and using the same imaging protocol, which has substantial implications for DTI research trials. Even when compared to a normal population imaged on the same MR scanner and using the same imaging protocol, substantial intra-scanner variability severely limits the ability to make a reliable diagnosis. The problem is further compounded when a normal control population is chosen from individuals imaged on different MR scanners and using different imaging protocols.

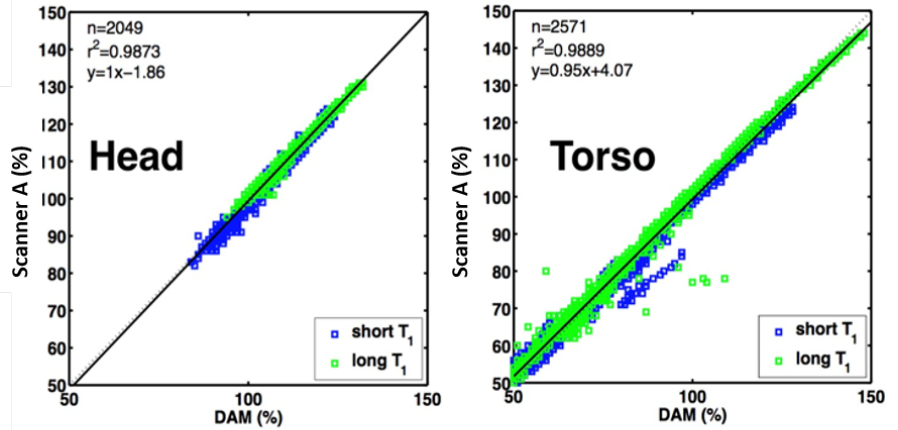
Evaluation of RF Transmit Calibration Options for Quantitative DCE-MRI, PI: Krishna Nayak, PhD, University of Southern California – Round 6 (2016-18 (NCE)) Project Z

The primary goals of this groundwork project were (1) to measure the agreement between the reference Double-Angle B1+ Mapping method (DAM) and the vendor-provide B1+ mapping options for different scanners, vendors, and sites, and (2) to use the measured B1+ errors to determine their impact on bias and variance of tracer kinetic parameters derived from DCE-MRI. This first required development of head-sized phantom (Diffusion Phantom Shell, High Precision Devices, Boulder, Colorado) and torso-sized (Shelley Medical, Hamilton, Ontario) phantoms with conductivity and T1 similar to tissue.



Torso- and head-sized B1+ mapping phantoms with conductivity and T1 relaxivity properties designed for this groundwork project.

	[NiCl ₂]	[NaCl]	Desired T1 (at 3T)	Desired Conductivity
Brain short T1	4.844 mM	40 mM	300 ms	0.6 S/m
Brain long T1	0.543 mM	40 mM	1500 ms	0.6 S/m
Torso short T1	4.844 mM	40 mM	300 ms	0.6 S/m
Torso long T1	0.543 mM	40 mM	1500 ms	0.6 S/m



Representative dataset showing good agreement between vendor-provided B1+ mapping and reference DAM mapping for one platform (Scanner A 3T).

We have computed the error (bias and variance) of all B1+ mapping methods compared to the (slow) DAM reference and will include detailed results in forthcoming publications. As expected, the amount of B1+ variation was greater in the torso phantom than in the head phantom and was greater at 3T compared to 1.5T for both phantoms. The spatial pattern of variation was comparable on all of the eight commercial 3T systems tested.

The short-term repeatability study requires a second round of data collection and is scheduled to happen within the few months for all scanners. Modeling and error propagation analysis is being performed as follows: We compute the expected error (bias and variance) of derived DCE-MRI metrics if performed without B1+ calibration information, or if performed with inaccurate B1+ calibration information. We use standard DCE-MRI simulation and error propagation analysis with a realistic arterial input function. Native contrast concentration vs time curves are generated using the extended Tofts-Kety model, and different combinations of pharmacokinetic parameters K^{trans} , v_e and k_{ep} within the physiologic range (defined as the native parameters Pnat) at a time resolution of 0.5s. Results of this analysis will be submitted for publication in *Medical Physics*.

PET-Amyloid Biomarker Committee:

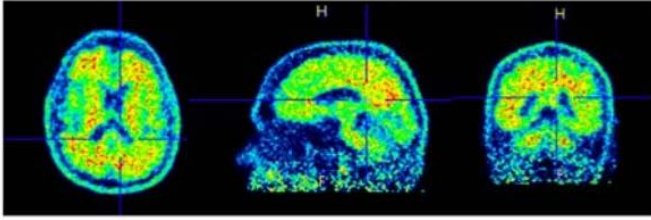
Quantification of Reconstruction Method Impact on Measured Amyloid Load. PI: Dawn Matthews, MS, MBA - ADM Diagnostics, LLC - Round-6 (2016-17) Project DD

This project quantified the impact of reconstruction method on brain amyloid measurement, and developed recommendations for reconstruction method and region of interest (ROI) definition based upon objectives for amyloid quantification. A comprehensive set of scans was systematically reconstructed using variations of algorithms and parameters (figure below). In addition to ROI analyses, a multivariate machine learning platform (NPAIRS) was used to characterize differences between reconstruction methods.

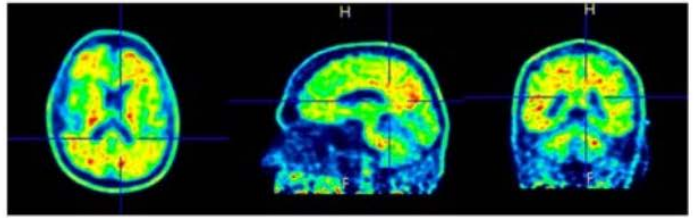
Key findings are illustrated below, where the choice of reconstruction method and parameters impacted amyloid SUVR. For a cortical average SUVR referenced to whole cerebellum, deviations were in the range of +/-4% from OSEM 4i24s 5mm values (used as a reference point) in the 15 florbetapir scans tested. Even this average variability, which is much less than the largest individual excursions observed and described below, suggests the importance of applying the same reconstruction method and parameters in longitudinal analysis.

QIBA Final Report - March 2018

OSEM 4i24s 0mm



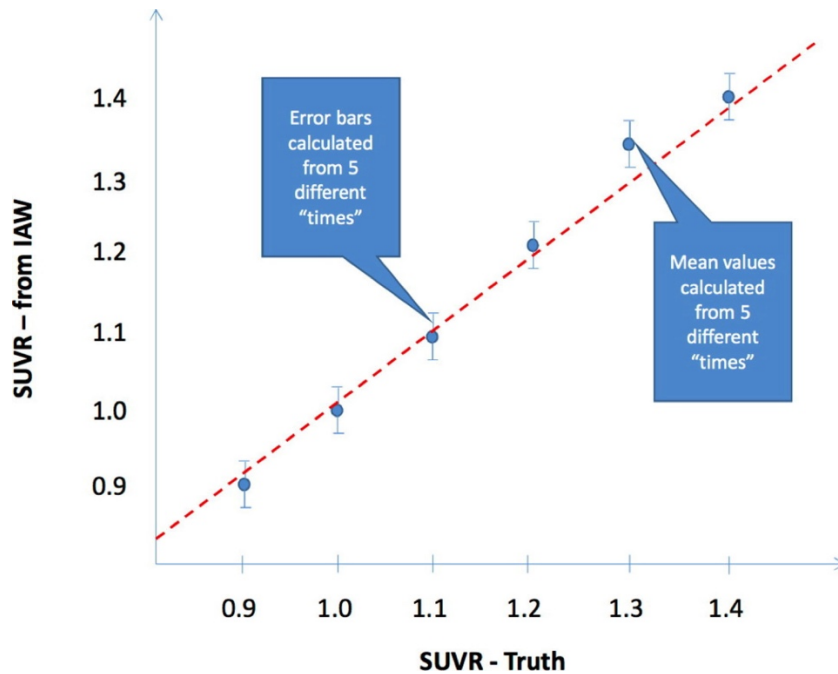
OSEM+PSF 4i24s 0mm



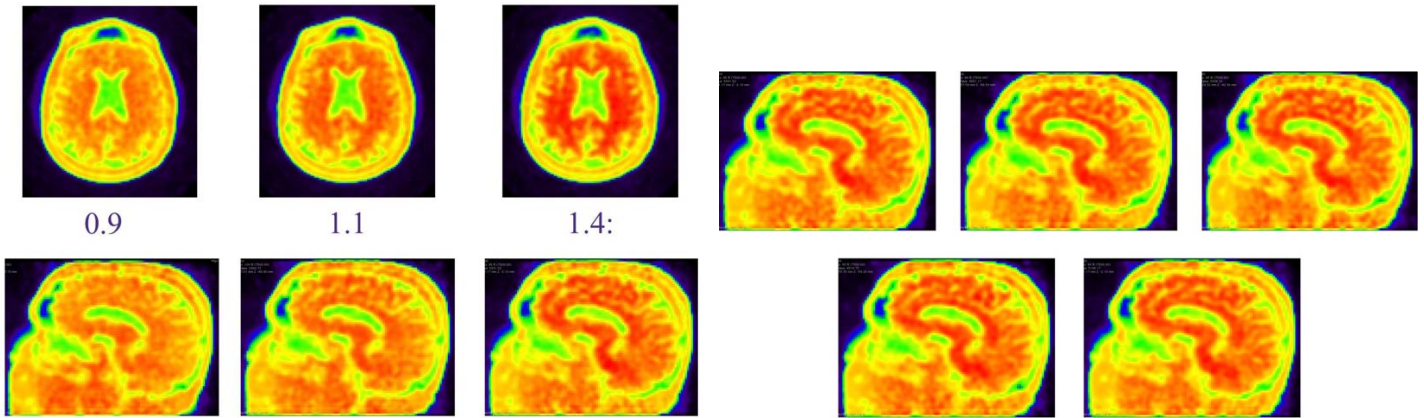
Example comparison of images generated using different reconstruction methods. Parameters were also varied.

Digital Amyloid Phantom for Software and Scanner Validation, PI: Paul Kinahan, PhD - University of Washington - Round-6 (2016-17) Project AA

This project's goal is the creation of a digital amyloid phantom to support efforts to better characterize the quantitative measurement of amyloid imaging agents for PET. This is the second phase of a project where prototype digital and physical phantoms were constructed. Many suggestions for improvements were listed in the final report for the first project. In this second phase, lessons learned from the first project were extended in an implementation of a series of amyloid DROs simulating a range of anatomical variants, with an array of amyloid distributions.



Intended approach with five realizations of six gray:white matter uptake ratios.

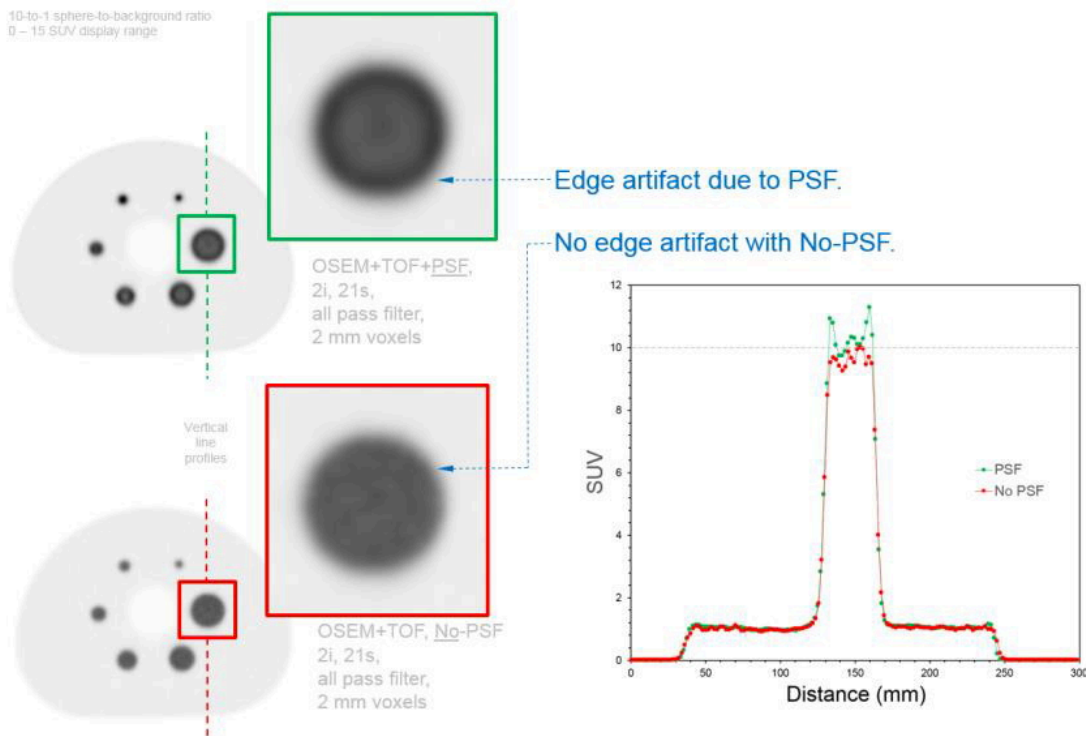


Left: Axial and Sagittal views from three of the six uptake ratios as indicated. Right: Showing 5 noise realizations from the 1:1.4 uptake ratio dataset.

FDG-PET/CT Biomarker Committee:

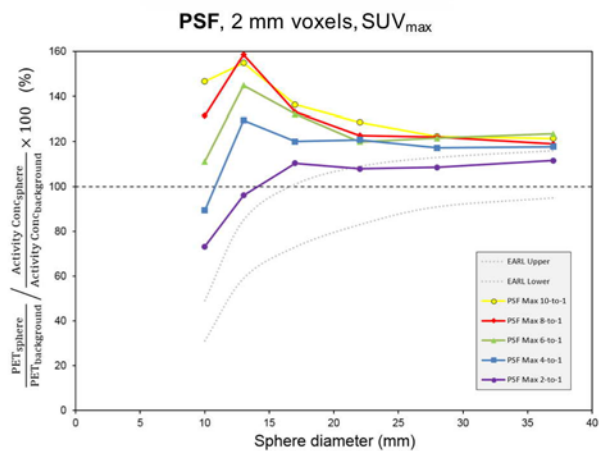
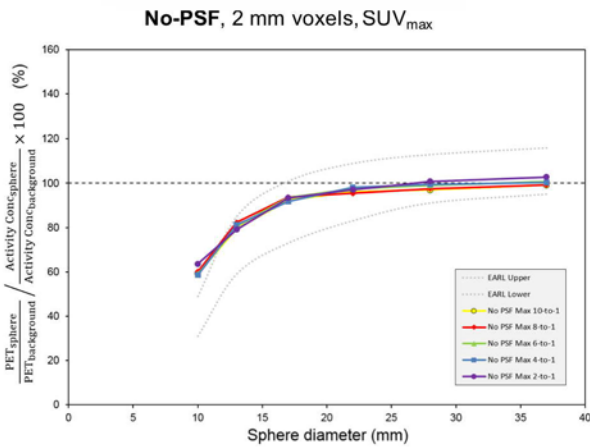
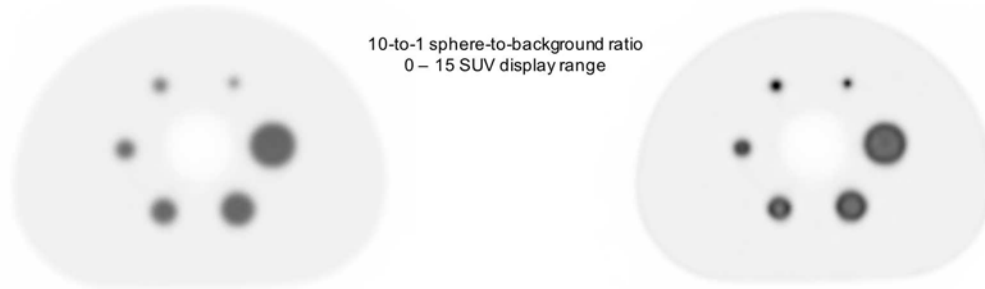
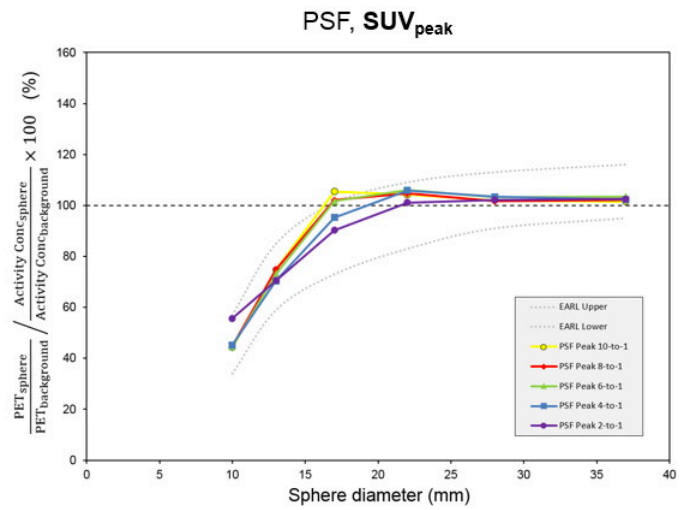
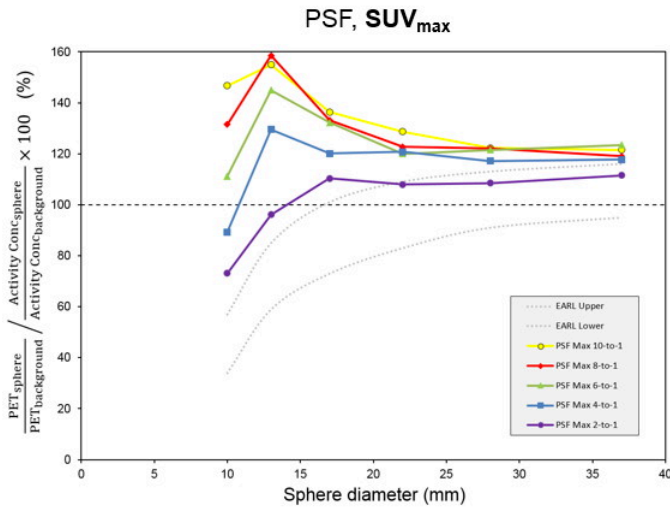
SUV Quantification with Point Spread Function PET Reconstruction, PIs: Martin Lodge, PhD, Johns Hopkins University and Ronald Boellaard, PhD, University of Groningen - Round-6 (2016-17) Projects R1 and R2

The aim of this study is to investigate the effect of PET reconstruction with Point Spread Function (PSF) modeling on tumor standardized uptake value (SUV). Initial work has involved phantom experiments using the NEMA image quality phantom, investigating the effect of image noise, voxel dimensions and phantom positioning on various SUV metrics including SUVmax, SUVpeak and SUVmean. It is now clear that PET reconstruction with PSF modeling, which is becoming common, significantly biases PET SUVs (figure below).



QIBA Final Report - March 2018

Results (below) suggest the SUV_{peak} may be preferable to SUV_{max} when using PSF modelling, at least for tumors greater than around 17 mm diameter. SUV_{peak} improves the quantitative characteristics of PSF images, reducing positive bias and decreasing the variability of tumor SUVs.



mCT, OSEM+TOF, 2i, 21s, 5 mm Gaussian, 400x400

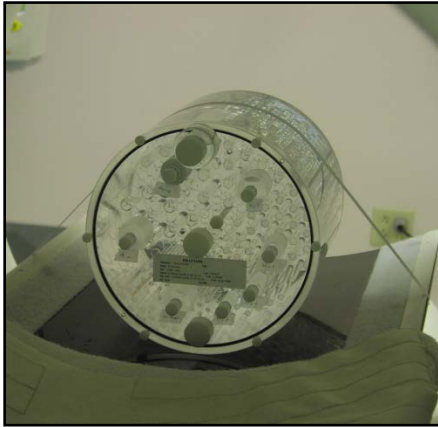
mCT, OSEM+TOF+PSF, 2i, 21s, all pass filter, 400x400

Simple Variable Estimates in PET, PI: Tim Turkington, PhD-Duke Univ-Round-6 (2016-18 (NCE)) Project CC

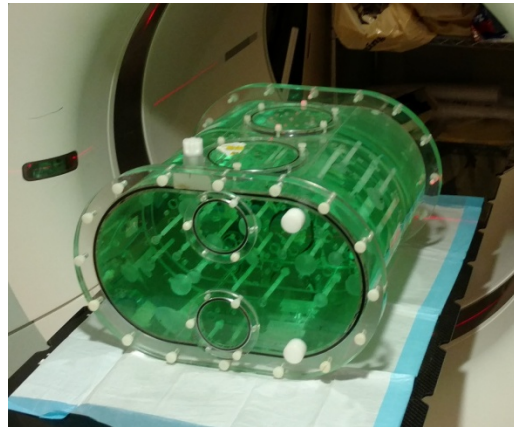
This project developed and tested methodology for assessing the scanner noise component of test-retest variability so future versions of PET Profiles can establish more specific and meaningful requirements for PET data acquisition and processing for individual imaging sites. A key missing feature in the current FDG-PET Profile is a specification on image noise as it pertains to test-retest of SUV in hot lesions. Most, if not all, PET/CT systems built in the last 15 years are capable of producing PET images that support the Profile claim. However, the scan parameters necessary for each system are certain to be different. For example, a system with lower sensitivity or rejection of background will require a longer scan to achieve the same quality. Various

QIBA Final Report - March 2018

benchmarks for image quality exist, but none is used routinely to assess test-retest variability. Many sites in the US are accredited by the ACR and therefore routinely image the ACR PET phantom. Repeat scans of this phantom versus an extended whole-body phantom were used as the basis for a variability assessment.

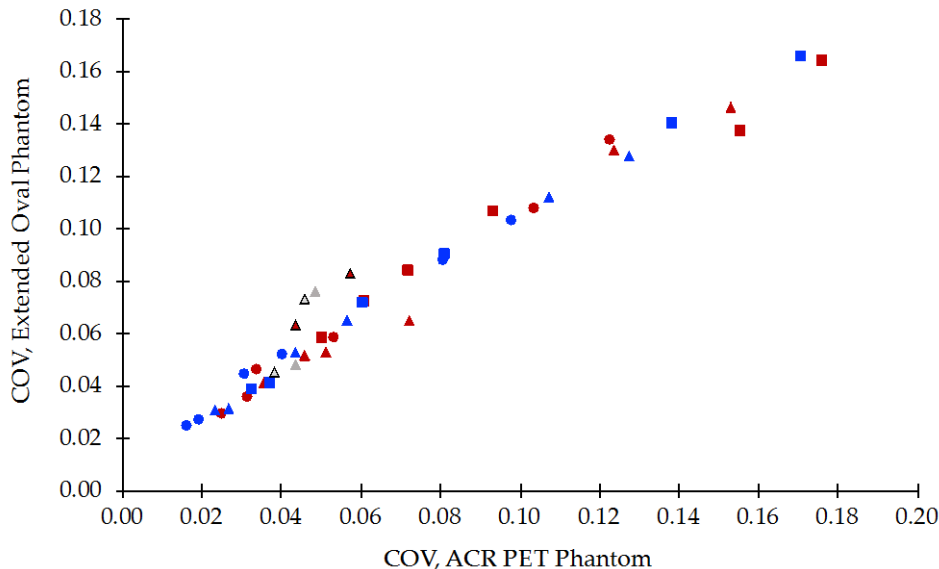


ACR PET (Esser) phantom



Extended PET whole-body phantom

The graph below demonstrates the strong correlation between coefficient of variation (COV) measurements in the whole-body phantom (3:1 contrast) vs. ACR phantoms. The 3-cm spheres in the whole-body phantom were compared with the 25-mm cylinder in the ACR phantom and 2-cm spheres were compared with the 18-mm cylinder. A slope of 1.0 was determined, showing that the variability measured in repeat scans of the ACR phantom are indicative of variability in a larger phantom (or average patient) in 2-3 cm lesions.



In conclusion it is feasible to use variability measurements from the ACR phantom to predict variability in larger bodies in 2-3 cm spheres (the smallest size that can accurately be quantified). To keep the image noise from dominating the overall variability in PET images, which has been shown to be ~12% in controlled studies, the image noise should be <6% in an average size patient. These results will be included in the requirements and conformance sections of the next version of FDG-PET Profile.

Biologic and Reader Repeatability of FDG and CT Volumetric Parameters (ACRIN 6678 & MERCK), PI: Rathana Subramaniam, MD, PhD – UT Southwestern Medical Center (FGD-PET/CT Biomarker Committee) – Round-5 (2015-2018 (NCE)) Project B

QIBA Final Report - March 2018

Three overall objectives of this project were to (1) establish the biologic repeatability of FDG PET metabolic tumor volume (MTV) and total lesion glycolysis (TLG) using the arm C data of ACRIN 6678 and MERCK data, (2) establish the inter reader reproducibility and variability of MTV and TLG using arm C data of ACRIN 6678 and MERCK data, and (3) establish the reproducibility using digital reference object (DRO) for volumetric parameters (MTV and TLG).

Summary of Results: Ninety-six patients were accrued at 17 sites for the ACRIN 6678 trial. Of these, 45 (recruited at 10 sites) consented to participate in the evaluation of test–retest repeatability; evaluable data are available for 34 of these patients. Merck provided data from 47 patients who were accrued at 14 centers in Europe and Asia from February 2009 to May 2010. Evaluable data are available for 40 of these patients.

The quantitative parameters are presented below as mean, SD and 95% CI of SD.

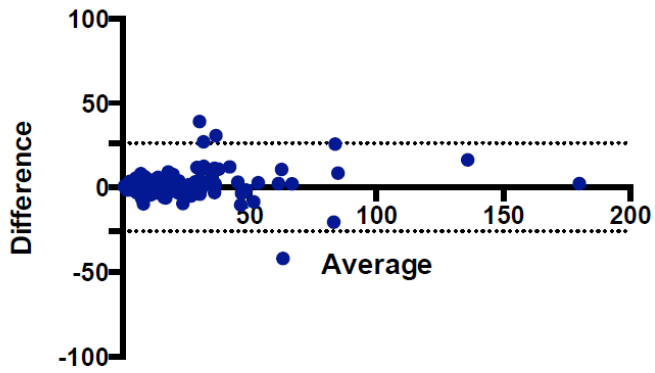
Parameter	Mean	SD	95% CI
MTV40 ₁	17.74	27.14	13.86 – 21.63
MTV40 ₂	17.98	23.24	14.65 – 21.30
MTV50 ₁	12.20	20.11	9.32 – 15.08
MTV50 ₂	11.88	16.82	9.48 – 14.29
MTVedge ₁	19.80	33.02	15.08 – 24.53
MTVedge ₂	20.85	34.32	15.92 – 25.77
TLG40 ₁	122.2	279.0	82.31 – 162.2
TLG40 ₂	123.0	285.7	82.16 – 163.9
TLG50 ₁	97.16	235.7	63.43 – 130.9
TLG50 ₂	94.98	242.3	60.31 – 129.7
TLGedge ₁	134.9	325.0	88.42 – 181.4
TLGedge ₂	138.7	343.4	89.42 – 188.0

There was excellent correlation between the baseline and repeatable measurements of MTV and TLG, as outlined below.

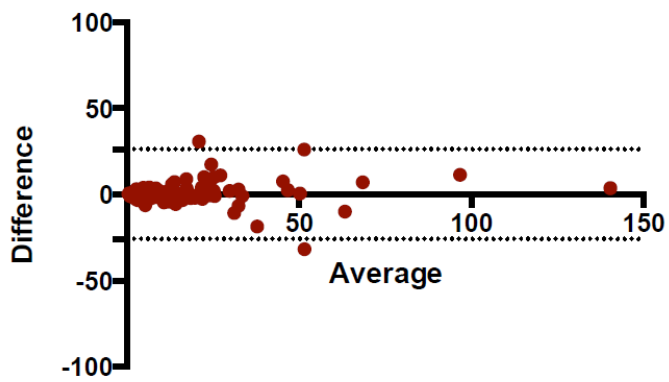
Parameter	Correlation Coefficient (r²)	P value
MTV40 ₁ - MTV40 ₂	0.943	< 0.0001
MTV50 ₁ - MTV50 ₂	0.944	< 0.0001
MTVedge ₁ - MTVedge ₂	0.942	< 0.0001
TLG40 ₁ - TLG40 ₂	0.977	< 0.0001
TLG50 ₁ - TLG50 ₂	0.975	< 0.0001
TLGedge ₁ - TLGedge ₂	0.968	< 0.0001

Bland – Altman analysis revealed there was no systematic increase or decrease from the first (baseline) to the second (repeated) PET/CT scan for MTV and TLG parameters for the target lesions.

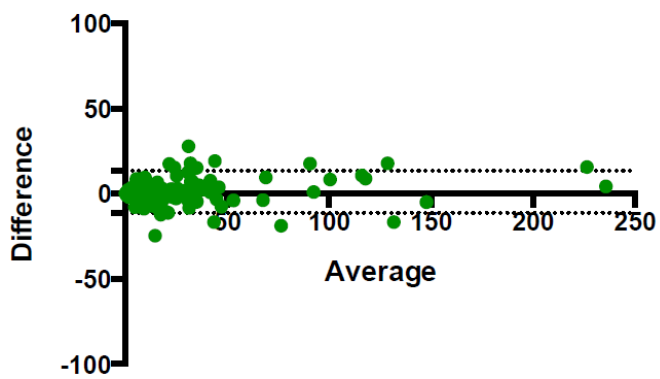
Bland-Altman MTV40



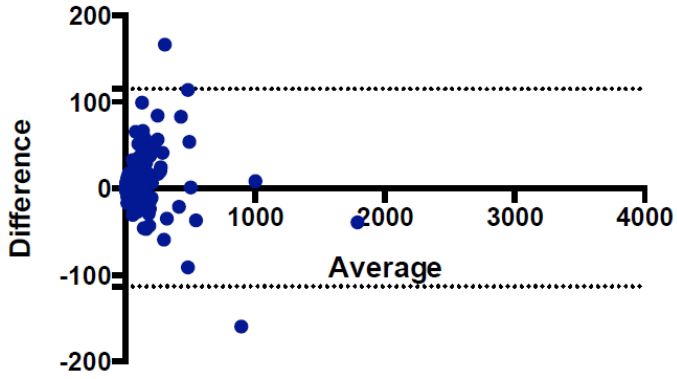
Bland-Altman MTV50



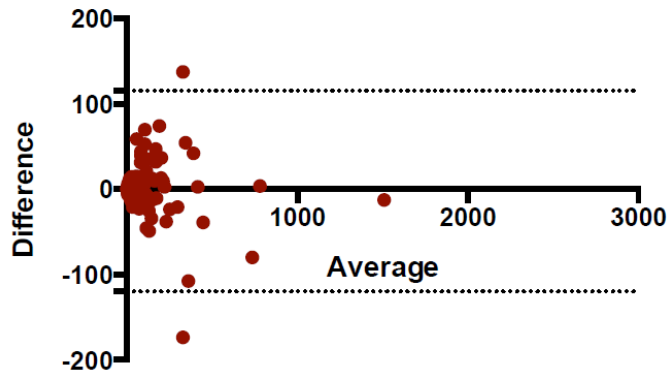
Bland-Altman MTVe



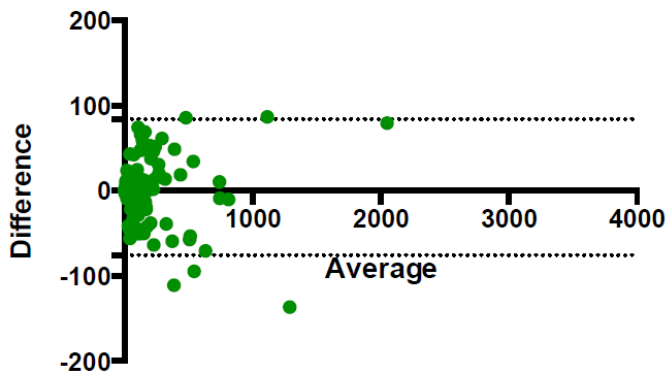
Bland-Altman TLG40



Bland-Altman TLG50



Bland-Altman TLGe



QIBA Final Report - March 2018

Parameter	Bias	SD	Upper 95% CI	Lower 95% CI
MTV40	0.23	13.29	26.29	-25.82
MTV50	-0.32	12.67	24.52	-25.15
MTVedge	1.13	6.3	13.49	-11.24
TLG40	0.81	58.21	114.9	-113.3
TLG50	-2.18	60.0	115.4	-119.7
TLGedge	4.28	40.73	84.11	-75.56

Conclusion: MTV and TLG, baseline and repeated measurements demonstrates excellent correlation coefficient (0.942 – 0.977) and with a bias between -0.32 to 1.13ml for MTV and -2.18 to 4.28 for TLG for the threshold and gradient method of segmentations.

Next steps: (1) More detailed statistical analysis (by ACRIN statistical staff), including investigating (obtaining data from ACRIN) the impact of uptake time and other factors on MTV and TLG repeatability and analyzing the ACRIN and MERCK cohorts, separately, and (2) Final manuscript submission (Dr. Subramaniam plans to continue analysis and manuscript preparation independently after this contract expires).

SPECT Biomarker Committee:

Multi-center Phantom Study to Characterize Bias and Precision of Quantitative I-123 SPECT. PIs: Yuni Dewaraja, PhD, University of Michigan and John Dickson, PhD, University College London Hospital - Round-6 (2016-17) Projects V1 and V2

The overall objective of this project was to determine the acquisition parameters and reconstruction methods for estimating the specific binding ratio (SBR) in 123-I ioflupane SPECT with higher precision and reduced bias.

A striatal phantom filled with the ratios suggested in the Profile was used to evaluate bias (relative to true SBR) and variance in SBR for different reconstruction methods with and without uniform (Chang) and non-uniform (CT-based) attenuation correction. Five acquisitions at 3 million counts each were performed on a SPECT/CT scanner (Scanner B) with the low energy collimator, and each data set was reconstructed using 5 methods. No post-filtering was used with OSEM, as recommended in the Profile when focus is on quantification.

Summary of Results: The pooled results are shown in the table below. For both scanners A and B, 1 million acquired counts is sufficient for SBR quantification to achieve sufficient precision. Bias in SBR of striata ranged from 47% (OSEM-ACSCRR) to 77% (OSEM- NC). The relative standard deviation (coefficient of variation) of the SBR ranged from 14.2% (OSEM- ACSCRR) to 3.0% (OSEM-NC). Bias in Caudate-to-Putamen ratio ranged from 18% (OSEM-ACSCRR) to 33% (FBP). Lower bias than SBR because partial volume effects are similar for the two striata. The relative standard deviation (coefficient of variation) of this ratio ranged from 7.7% (OSEM-ACSCRR) to 1.0% (FBP).

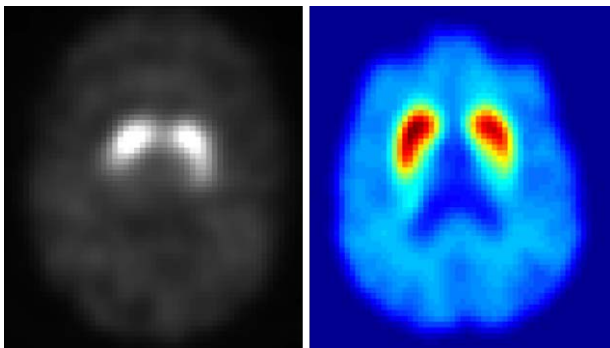
QIBA Final Report - March 2018

Scanner	Recon	R. Puta SBR	R. Caud SBR	L. Caud SBR	L. Puta SBR	L Caud/Put Ratio
A	ACSCRR	46.71 (8.39)	64.70 (10.51)	63.12 (8.89)	50.19 (14.16)	25.60 (7.68)
	NC	65.19 (4.40)	76.62 (4.68)	75.50 (9.81)	61.98 (7.37)	32.94 (4.54)
	FBP	64.80 (4.31)	75.08 (3.36)	75.29 (5.84)	61.20 (4.23)	33.25 (3.47)
B	ACSCRR	56.51 (4.23)	65.76 (6.03)	60.85 (4.81)	55.12 (5.03)	18.44 (1.25)
	NC	66.72 (2.95)	73.42 (4.13)	72.10 (3.85)	64.85 (5.11)	24.82 (1.00)
	FBP	60.59 (3.38)	67.86 (6.65)	66.26 (4.49)	59.23 (4.42)	21.84 (1.42)

Summary of pooled results for A and B scanners. Percent Bias in Specific Binding Ratios for the four striatal components and diseased Caudate-to-Putamen Ratio for OSEM with corrections (ACSCRR), OSEM with no corrections (NC) and FBP with no corrections. Relative standard deviations (%) are shown in parentheses. Results are for 5 repeat acquisitions with LE collimator. OSEM is for ~ 100 equivalent iterations and with post-filtering.

I-123 DAT Scan Digital Reference Object Development. PI: Robert Miyaoka, PhD - University of Washington - Round-6 (2016-18 (NCE)) Project X

The primary goal is to design and construct a prototype brain Digital Reference Object (DRO) phantom with properties appropriate for testing software used to characterize I-123 ioflupane uptake in the striatum and the derivation of striatal specific binding ratio in a quantitative fashion. The results can be used to develop methods to be used in the SPECT DAT scan Profile and to test I-123 DAT scan analysis software in a consistent fashion (figure below). Based upon the initial results from other QIBA committees that have developed DRO phantoms, it is clear that consistent testing of vendor data analysis software is necessary to validate/confirm that given the same reference image all vendor software will return the same values given the same analysis task.



Left: Clinical DAT scan. Right: DRO after noise, smoothing, and down-sampling have been applied.

The DRO was sent to a number of sites, including members of the DAT SPECT task force (*i.e.*, Drs. John Seibyl, Yuni Dewaraja, and John Dickson) and two investigators in Japan (*i.e.*, Drs. Iida and Nakahara). In total, the DAT DRO was processed by 6 sites; four different analysis packages with one analysis package used three times on the images. The results shown below demonstrate the variation between the 4 different vendor analysis packages as well as variation of the same analysis package used at 3 different sites. These methods will be used to define criteria for the next version of the Profile.

QIBA Final Report - March 2018

Analysis Software	Striatum SBR		Caudate SBR		Putamen SBR	
	Right	Left	Right	Left	Right	Left
Truth	4.5	4.5	4.5	4.5	4.5	2.25
Vendor 1 (no blur)	2.9	2.05	3.36	2.76	2.7	1.69
Vendor 2 (no blur)	3.19	1.87				
Vendor 3 (no blur)	2.53	1.81	2.4	2.33	2.56	1.43
Vendor 4 (no blur)			3.6	3.05	2.76	1.7
Vendor 1 (6 mm blur)	2.59	1.8	3.02	2.34	2.4	1.52
Vendor 2 (6 mm blur)	3.23	1.86				
Vendor 3 (6 mm blur)	2.13	1.53	2.3	1.99	2.05	1.2
Vendor 4 (6mm blur)			3.19	2.61	2.2	1.39
Vendor 1 (10 mm blur)	2.17	1.52	2.57	1.9	2	1.32
Vendor 2 (10 mm blur)	3.23	1.86				
Vendor 3 (10 mm blur)	1.72	1.21	1.92	1.61	1.66	0.91
Vendor 4 (10 mm blur)			2.67	2.11	1.71	1.12

Table. DAT DRO SBR results for the same analysis program at 3 sites.

Analysis Software	Striatum SBR		Caudate SBR		Putamen SBR	
	Right	Left	Right	Left	Right	Left
Truth	4.5	4.5	4.5	4.5	4.5	2.25
Vendor 1a (no blur)	2.9	2.05	3.36	2.76	2.7	1.69
Vendor 1b (no blur)			3.35	2.75	2.58	1.67
Vendor 1c (no blur)	2.79	2.02	3.29	2.7	2.58	1.67
Vendor 1a (6 mm blur)	2.59	1.8	3.02	2.34	2.4	1.52
Vendor 1b (6 mm blur)			2.99	2.33	2.39	1.5
Vendor 1c (6 mm blur)	2.55	1.79	3.01	2.35	2.36	1.51
Vendor 1a (10 mm blur)	2.17	1.52	2.57	1.9	2	1.32
Vendor 1b (10 mm blur)			2.54	1.89	1.98	1.3
Vendor 1c (10 mm blur)	2.15	1.5	2.54	1.89	1.98	1.3

US SWS Biomarker Committee:

Variances between measurements on ultrasound systems by 8 vendors at 13 sites on 9 commercial systems with 2 transducers each have been analyzed in 3 each of elastic and viscoelastic phantoms. The initial plan to validate system measurements by comparison with MRE results was changed due to the high variance levels permitted in the first iteration of the MRE Profile. Individual system results will be validated by comparison with the values obtained in phantoms on a well-specified research ultrasound system and algorithm. Rather than a purely experimental correction for frequency, a model-based correction is being sought in simulations that might be more universally correct *in vivo*. This correction is being studied for a later second, publically-distributed Profile. Conformance will be monitored by completion of checklists and a new plan for conformance of each actor with expected values of variability for that actor.

Establishing Acceptable Variance Limits for Healthy, F1 and \geq F2 Fibrosis Shear Wave Speed Values Across Systems and Between Operators for the QIBA Profile, PI: Manish Dhyani, MD - Mass Gen Hospital - Round-6 (2016-18 (NCE)) Project BB

Analysis was performed on 23 studies completed, as of 3/9/2018, of the planned 30 subjects at varying liver fibrosis stages. These studies at Massachusetts General Hospital included repeated measurements with up to 5 different US systems. The Washington DC VA Medical Center will study the remaining patients this month. The studies were performed to: 1) Characterize intraobserver and interobserver and intersystem variability of liver SWE in five commercially available systems using SWS measurements; and 2) Determine the effects on

QIBA Final Report - March 2018

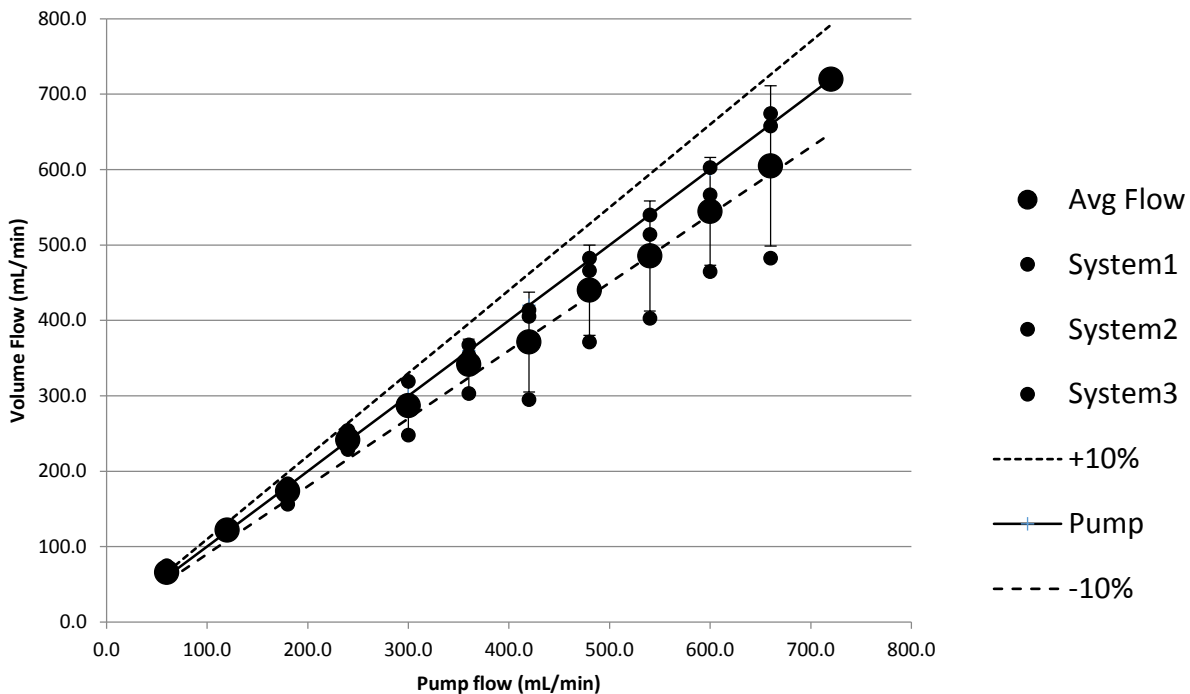
SWS of a subset of possible acquisition errors. Additionally, numerous meeting presentations with abstracts and one poster were presented this period.

Analysis of Sources of US SWS Measurement Inter-System Variability, PI: Mark Palmeri, MD, PhD Round-5 (2015-16) Projects F1 and F2 produced one publication and one submitted in *IEEE Trans UFFC*, as well as at least 5 meeting presentations and one proceedings article.

US Volume Blood Flow (VBF) Biomarker Committee:

Examination of Flow Phantom as Reference Standard for Validation of Ultrasound Volume Blood Flow Measurement, PI: Oliver Kripfgans, PhD - University of Michigan - Round-6 (2016-18 (NCE)) Project W

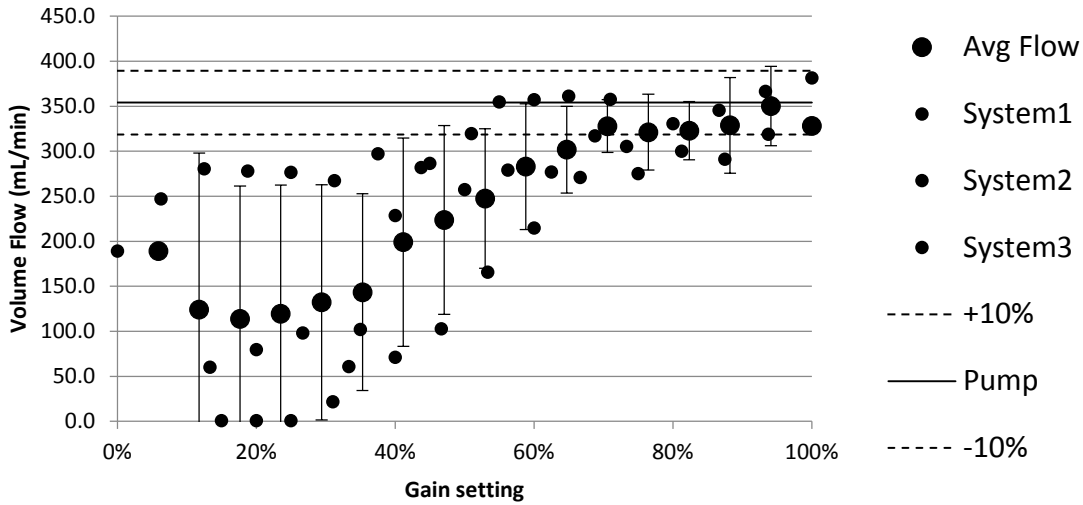
This project has now included all nine out of nine system/site combinations of field tests with phantoms as described in other sections. With the support of QIBA/NIBIB funding, Dr. Pinter has traveled to five participating sites (University of Washington, University of Alabama at Birmingham, the Mayo Clinic in Minnesota, the University of Wisconsin and Rocky Vista University in Utah) to assist with system setup and training for data acquisition. The team is engaged in telephone and regular video conferencing with five participating manufacturers to enable access to data essential for volume flow computation on six different platforms. Testing of the groundwork protocol was performed on three separate company platforms on a prototype groundwork flow phantom developed in cooperation with an established ultrasound phantom company (Gammex Inc.). The average coefficient of variation is 4.6%. Two of the three systems tested with full analysis exhibit a mean +2% bias and one ~-35% bias. The latter is due to a simple, correctable error in data output from the system. Linear frequency shift and Doppler power image data extraction and partial volume corrections are in process with cooperation from two additional companies. Two phantoms (one supported by a professional/research society) were shared for testing the groundwork protocol at 6 academic sites and 1 industry site. To date more than 18,450 volume datasets have been collected.



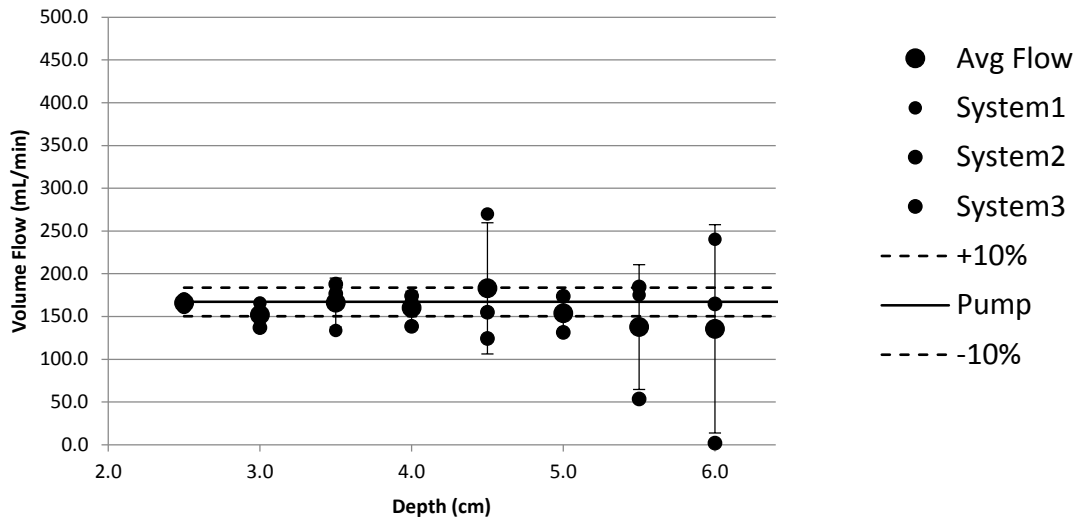
Results for assessment of constant blood volume flow in a realistic flow phantom, *i.e.*, a phantom with non-uniform tube diameter and curved tubing path. Three systems tested at three sites. One system is reading low at flows greater than 400 mL/min, which is being investigated. Other systems yielded

QIBA Final Report - March 2018

values within $\pm 10\%$, an excepted range. Large markers are the average of three systems. Small markers are the individual anonymized systems.

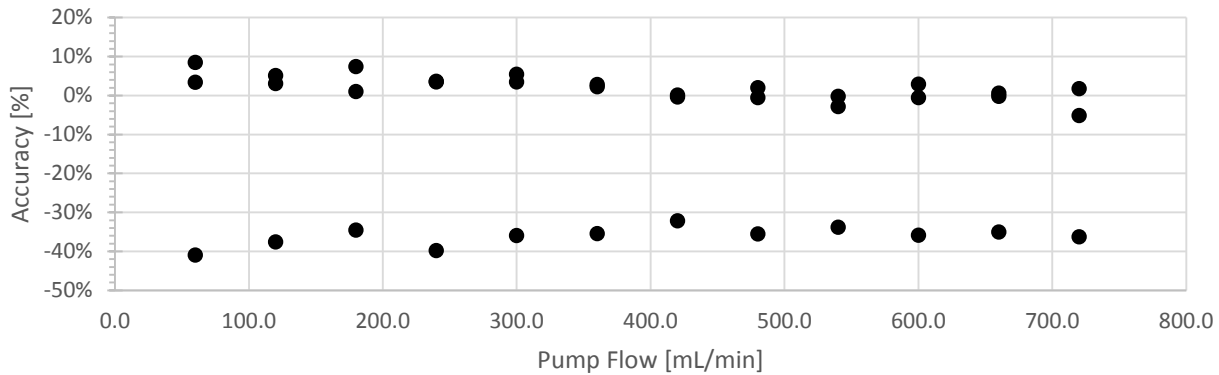


Results for assessment of constant blood volume flow of 350 ml/min as a function of receive gain setting. Three systems tested at three sites. While there is large variability in the system responses at low gain, as is expected, for system gains greater than approximately 70%, all systems converge towards true flow. Dotted lines represent $\pm 10\%$.



Results for assessment of pulsatile blood volume flow as a function of depth. The initial goal was to reach 5 cm with $\pm 10\%$ error, which is largely achieved. Deeper depth was investigated to determine current depth limitations for the selected transducers.

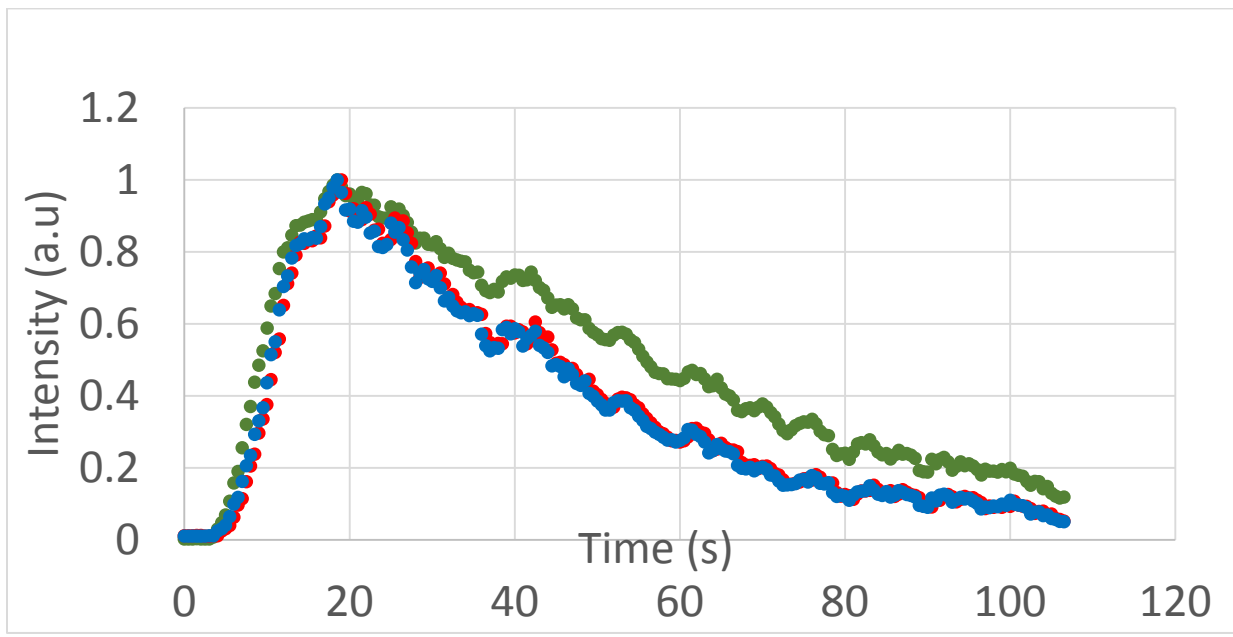
Accuracy of 3 Systems across 3 Sites



Results for accuracy across platforms and sites for one of the eight objectives, variation in flow, for non-pulsatile flow. Two systems show an approximate positive bias of 0 to 10%. Across the entire flow range this bias averages to 2% high. Another system is underestimating and we are in communication with the manufacturer to remedy this situation.

CEUS Biomarker Committee:

Groundwork is progressing with a standard phantom in one laboratory and parts for these phantoms have been distributed to two other laboratories. Major international clinical and basic research groups and numerous manufacturers are participating in literature analysis and planning meetings and testing. Some standardization of measurements has already occurred as shown in the following figure.



Time intensity curves from the same data set with ultrasound contrast in the QIBA CEUS flow phantom. They are processed with 3 different quantification software packages demonstrating differences, later corrected on commercial software.

Cross-Modality:

US SWS and MRE Biomarker Committees:

The Round-5 (2015-16) Project F1 and F2 teams performed further analysis of MRE and US SWS data on phantoms to show the level of corrections needed to provide the same results from the two modalities in the elastic and viscoelastic phantoms.

Aggregated Measures of Agreement for QIB Validation: An Open Source Toolkit, PI: Daniel Barboriak, MD - Duke University - Round-5 (2015-16) Project K

The purpose of this project was to develop open-source software to calculate aggregated measures of agreement in order to facilitate image analysis algorithm development, comparative analysis of algorithm output, and demonstrate technical conformance. This project also developed a toolkit that can calculate the following statistics: concordance correlation coefficient, root mean square deviation, total deviation index, Bland-Altman limits of agreement, and Sigma analysis based on estimates of allowable total error. An existing open-source package, QIBA DRO Evaluation Tool (QDET), developed by Dr. Hendrik Laue in a previous groundwork project, was used as a starting point. This package's source code was downloaded, and preliminary analysis of it has been conducted. The software required some Python modules, such as WxPython and Matplotlib, which are not part of the standard Python installation. These additional modules were obtained and installed. Python has several ways to package and release software as an executable application; these were investigated to determine which to use.

We have successfully continued updating the capabilities of the QDET program. The program now accepts text as input, and we have validated that the RMSD, CCC, TDI and Bland-Altman statistics obtained from text images are identical to those obtained from corresponding image data. These statistics have been verified against statistics obtained from the R software package. Scripts demonstrating how to use QDET to tune software parameters and to rank performance of competing algorithms have been completed.

OBJECTIVE 4. DEVELOP AND EMPLOY PHYSICAL AND/OR VIRTUAL (DIGITAL) REFERENCE OBJECTS NEEDED FOR ASSESSMENT OF IMAGING BIOMARKER VARIABILITY AND/OR TO DEMONSTRATE COMPLIANCE WITH QIBA PROFILES.

Selected specifics:

CT Volumetry Biomarker Committee:

Reference Image Set for Quantitation Conformance of Algorithmic Lesion Characterization, PI: Ehsan Samei, PhD - Duke University (CT Volumetry Biomarker Committee) - Round-5 (2015-16) Project C

A database composed of 100 patient thoracic images and 100 patient abdominal images was collected from Duke University Hospital. Using the Duke database of lesion masks and their cloned renditions, 80 lesion morphologies (ground glass, part-solid, and solid with varying shapes and contrast and with sizes ranging within 4-20 mm) were virtually inserted into each patient lung image set. All lesions contain realistic internal heterogeneity. Similar insertions were applied to liver images. Patient images and lesion models were examined by experienced radiologists. This database has been uploaded to the QIDW.

PDF-MRI Biomarker Committee:

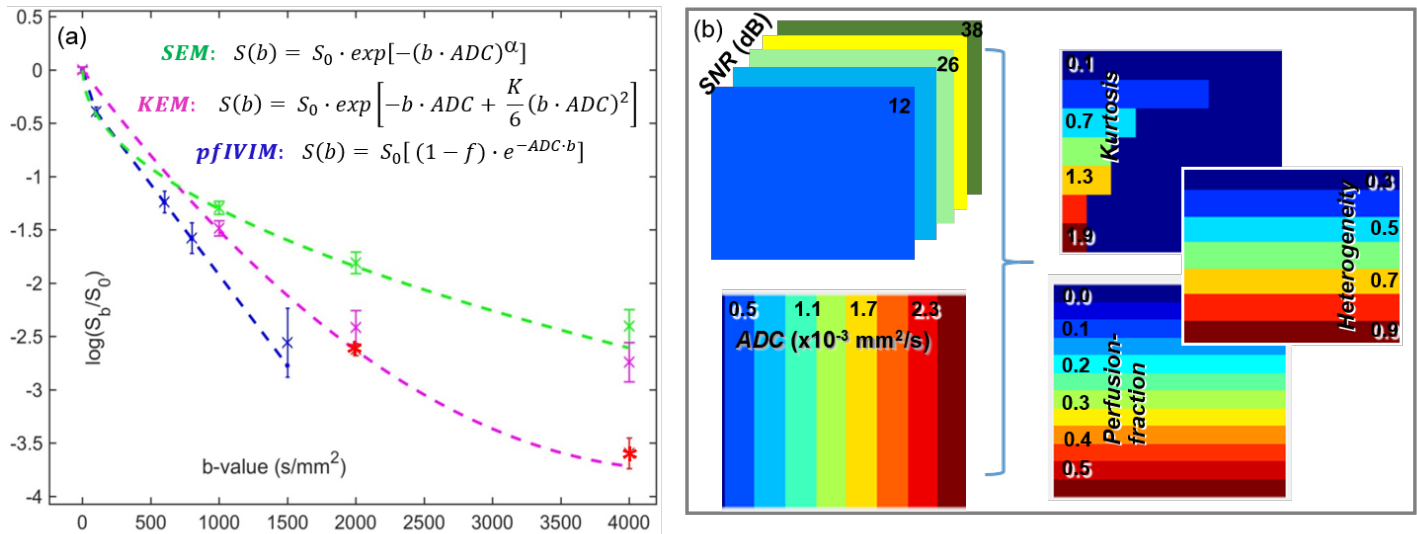
The DWI-MRI Task Force has designed, built, and circulated for system testing its PVP-based, thermally-controlled-by-ice-water physical DWI phantom with associated analysis software (Round-3 Projects A-B). This phantom is now commercially available from High Precision Devices Inc. (Boulder, CO), and has been recently incorporated in multi-center clinical trials.

DWI-DRO Development for ADC Analysis, PI: Dariya Malyarenko, PhD – University of Michigan (PDF-MRI Biomarker Committee / DWI Task Force), Round-5 (2015-16) Project G.

Development of this DWI-DRO is complete and it is suitable to assess robustness of DWI analysis software packages used for quantitative ADC map generation. Modeled tissue properties span $ADC = 0.1$ to 3.5×10^{-6}

QIBA Final Report - March 2018

$^3 \text{ mm}^2/\text{s}$ and $\text{SNR} = 0$ to 100 (of the $b=0 \text{ s}/\text{mm}^2$ image) within a DICOM-compliant DWI-DRO at b -values= $0, 500, 800, 2000 \text{ s}/\text{mm}^2$ (see Objective 3). Analysis of the DRO was used to predict bias and variation metrics over relevant ADC and SNR ranges (illustrated in figure below), which led to technical specifications stated in the DWI Profile. The DWI DRO was subsequently extended to non-Gaussian diffusion with the synthesized DICOM and its description available at: <https://www.researchgate.net/publication/322404473>.



DRO extended to non-Gaussian DWI: (a) Examples of log-ratio signal dependence on b -value for non-Gaussian (noiseless dashed) and $\text{SNR}_{\text{eff}}=31\text{dB}$ ($\text{SNR}_{b0}=26\text{dB}$, "x") diffusion models, color-coded according to analytical expression (legend). (b) DRO encompasses 3 DWI models: perfusion-fraction IVIM (pfIVIM); Kurtosis exponential model, (KEM); and stretched exponential model (SEM).

Dynamic Susceptibility Contrast MRI Phantom, PI: Ona Wu, PhD - Harvard University / Massachusetts General Hospital (PDF-MRI Biomarker Committee/DSC-MRI Task Force)-Round-5 (2015-18 (NCE)) Project H

Two phantom prototypes to estimate reproducibility across imaging sites were designed and fabricated. A high-level, generalized, cross-vendor imaging protocol usable on clinical MR scanners was also developed and utilized to characterize the DSC phantoms and assess repeatability on multiple scanner platforms. Phantom components have been tested for stability and suitable magnetic properties, such as susceptibility range and T_1 .

Evaluation of RF Transmit Calibration Options for Quantitative DCE-MRI, PI: Krishna Nayak, PhD, University of Southern California – Round 6 (2016-18 (NCE)) Project Z.

The purpose of this project was to measure and compare the spatial RF transmit inhomogeneity across different scanners, vendors, and sites, and to compare vendors' implementation of methods for RF transmit measurement that can be used to correct data prior to quantitative DCE-MRI analysis. For this, two phantom types were fabricated with fill-solution conductivity and relaxivity values that reasonably match those of human tissue: (1) an adult torso-sized and -shaped phantom and (2) a human head-sized phantom with a single interior fillable volume. These phantoms were scanned at multiple timepoints on 13 MRI systems located in the greater Los Angeles Area, within practical driving distance. These data will be used to (1) assess agreement between the reference (double-angle) B_1+ mapping method and the vendor-provided options for different scanners, vendors, and sites, and (2) use the measured B_1+ errors to determine their impact on the bias and variance of tracer kinetic parameters derived from DCE-MRI. Dr. Nayak plans to continue analyzing data independently after the Round-6 contract expires.

fMRI Biomarker Committee:

Three different rounds of fMRI DROs were developed. A Round-3 (2013-14) groundwork project involved generating 20 empirically-derived DROs, which were then downloaded and analyzed at 8 QIBA fMRI Biomarker Committee sites; comparing results from different sites helped develop analysis standards for Profile v1.0. A Round-4 (2014-15) project generated synthetic DROs to model one of three major forces of variance: head motion, neurovascular uncoupling, or variable task performance. A third (unfunded) round of

QIBA Final Report - March 2018

DRO development is ongoing to generate higher resolution DROs that combine empirically-derived head motion, neurovascular uncoupling, and task performance variability. Examples of Round-3 and -4 DROs are already available via the QIDW, and the next Round of DROs should be available in 2017 for use in field tests. These DROs will be used to synthesize subject datasets (N=20-30) representative of adequate quality to pass subject QC review, thus measured COM location should be within Profile claims.

PET-Amyloid Biomarker Committee:

A digital amyloid phantom (DRO) has been created to support efforts to characterize the quantitative measurement of amyloid imaging agents for PET using different analysis packages. The 30 images are comprised of 6 sets with different white matter to gray matter ratios (WM:GM) of 0.9, 1.0, 1.1, 1.2, 1.3, and 1.4. Each set has 5 noisy realizations. The 30 images were tested using a commercial analysis packages for PET amyloid tracer uptake measurement.

SPECT Biomarker Committee:

The specifications for a “fillable” phantom were developed for deployment in multi-center environments. A concept for a solid phantom, with a traceable source of Tellurium-123 as a proxy for Iodine-123, was developed in collaboration with subject matter experts from the National Institute of Standards and Technology. A Digital Reference Object (DRO) based on an anthropomorphic basal ganglia plus brain phantom has been developed based on the successful prototype by the FDG PET/CT group.

US SWS Biomarker Committee:

Analysis of Sources of US SWS Measurement Inter-System Variability, PI: Mark Palmeri, MD, PhD. Round-5 (2015-16) Projects F1 and F2 produced one publication and one submitted in IEEE Trans UFFC, as well as 3 meeting presentations and 2 posters. These are based on several rounds of elastic and viscoelastic phantom development, multisite tests thereof and commercialization.

The committee is developing specifications for construction of new viscoelastic phantoms to cover more of the range of viscoelastic moduli encountered in pathology. Purely elastic phantoms might still be used in the compliance tests.

US Volume Blood Flow (VBF) Biomarker Committee:

Examination of Flow Phantom as Reference Standard for Validation of Ultrasound Volume Blood Flow Measurement, PI: Oliver Kripfgans, PhD - University of Michigan - Round-6 (2016-18 (NCE)) Project W

Two phantoms with more realistically complex flow patterns were constructed after the prototype testing and have been delivered to the University of Michigan where they have been used for the round robin study. A groundwork protocol has been designed to test the variability of single vessel volume flow measurements using a new scanning pattern and algorithm for much lower variance and bias. Some preliminary results show variances between manufacturers and sources; these are being investigated. However, for this early stage of testing, the agreement among platforms is encouraging as is the engagement of the manufacturers in the implementation of volume flow. A new multimodality MRI/ultrasound flow phantom prototype is being tested.

CEUS Biomarker Committee:

Parts for three basic flow phantoms for CEUS studies were acquired at the University of Washington. One phantom was assembled there and the other two were shipped to co-leaders at The Surgical Hospital at Southwoods, Boardman, OH and the University of Texas at Dallas. The intent is to evaluate the little-studied variability of results between contrast agents and ultrasound systems.

OBJECTIVE 5. DEVELOP PROCEDURES AND PROCESSES FOR HARDWARE AND SOFTWARE MANUFACTURERS TO DEMONSTRATE CONFORMANCE WITH QIBA PROFILES.

Selected specifics:

CT Volumetry Biomarker Committee:

The procedures and processes have been revised, taking into account the feedback provided by the sites involved in the feasibility test.

CT Lung Density Biomarker Committee:

Committee members are actively working with vendors to develop models for harmonizing CT lung density measures across different scanner manufacturers and models. This effort primarily uses the COPDGene Phantom modified to contain NIST-certified foam standards that fall within the ranges expected for CT density of lung parenchyma. A first round of scans was performed by the vendors and a preliminary correction model was developed to achieve consistent performance within 2 Hounsfield Units across four vendor platforms. A second round of phantom scans with a greater number of scanner models and an updated anthropomorphic phantom was completed. Round-2 results support methods for harmonization of lung density measures across CT make and model using reduced dose (automated exposure control - AEC) and physical models developed in Round-1 with standardized reference materials (lung density equivalent foams). A Round-6 (2016-18 (NCE)) Project U is now testing these methods in an ongoing clinical research study, COPDGene. A conformance statement based on the Round-2 results is currently under development in collaboration with the vendor representatives on the committee.

PDF-MRI Biomarker Committee:

The conformance section of the DWI Profile has been completed according to the revised Profile template. Profile activities for the various “actors” are outlined in DWI Profile Section 3 along with assessment procedures in Section 4. Scanner platform-specific acquisition parameters, test procedures and target technical specifications are provided in Appendix E of the DWI Profile.

Analogous to other QIBA Profiles, such as DCE-MRI v1.0, conformance elements include scanner hardware, key personnel and image analysis procedures. The DTI and DSC efforts are fairly early in their Profile development, but conformance will be addressed as claims and protocols become more firm; scope and style will be consistent with the DWI and DCE Profiles.

fMRI Biomarker Committee:

The Round 1-5 groundwork projects on reproducibility and DROs identified reproducibility benchmarks for motor and language fMRI scanning that can be used as goals for conformance testing by manufacturers of fMRI task and analysis software. Currently there are no standards for such software, nor benchmarks for evaluating the quality of data produced by different tasks. Definitions and means to measure fMRI activation center of mass (COM), the target measurand, are not standardized and may not be readily available on standard fMRI analysis software packages, thus QIBA guidance and software tools may be required.

FDG-PET/CT Biomarker Committee:

The Profile has completed two rounds of feasibility-testing to examine its feasibility and practicality. The first round was at four academic imaging sites, while the second round was at a distribution of 12 regional clinical sites. At the completion of the first round feasibility test, representatives of each of the four vendors of PET/CT scanners systematically reviewed and commented upon their current and future ability to achieve conformance with the FDG-PET/CT Profile. Specific comments and suggestions from the vendors regarding feasibility and practicality were used to revise the Profile before the second round of feasibility-testing. After the second round of feasibility tests, checklists of conformance procedure steps for (1) manufacturers and (2) sites were added as appendices. These checklists have been submitted for publication as part of an overview paper and a revision is in process.

PET-Amyloid Biomarker Committee:

As noted above, the PET-Amyloid Profile is almost through the Public Comment phase; members of the Biomarker Committee have reviewed public comments.

SPECT Biomarker Committee:

A checklist for testing conformance has been developed as part of the Profile. Groundwork projects are being developed using a multi-center phantom study to characterize the bias and precision of quantitative I-123 SPECT. Specifically, this will allow characterization of the linearity of the confidence intervals surrounding precision and bias, which are suspected of increasing with disease severity, *i.e.*, increasing with decreasing signal-to-noise ratios as the signal becomes lost in patients with Parkinson's disease. In parallel, an I-123 DAT scan Digital Reference Object (DRO) has been developed. In addition to analyzing the phantom and DRO data, it is anticipated that the SPECT Biomarker Committee will assemble a test set from patients and matched controls with which all software vendors can test their analytical processes.

US SWS Biomarker Committee:

It is anticipated that phantom tests of shear wave speed estimation under various operating conditions will be an acceptable test for vendor conformance as claimed by the vendors. Round-6 (2016-18 (NCE)) Project BB will confirm or modify methods of archiving of acquired studies necessary for later site QA and/or conformance review that might also be useable for verification of ultrasound system performance in the field. A new round of testing by manufacturers using the visco-elastic phantoms under development is likely.

An example of the level of provided aid to validating conformance follows:

QIBA Final Report - March 2018

Actor	Activity	Section
Ultrasound Scanner (Acquisition Device)	Pre-delivery	3.1.
	Image Data Acquisition	3.6.
Scanner Vendor	Installation	3.2.
	Periodic QA	3.3.
Technologist/Sonographer	Subject Selection	3.4.
	Subject/Patient Handling	3.5.
	Image Data Acquisition	3.6.
	Image Data Reconstruction	3.7.
Radiologist	Subject Selection	3.4.
	Subject/Patient Handling	3.5.
	Image QA	3.8.
	Image Analysis	3.10.
QA Manager (QA Specialist, Medical Physicist or Hospital Engineer)	Acceptance Testing	3.3.
	Periodic QA	3.3.
Reconstruction Software	Image Data Reconstruction	3.7.
Image Analysis Tool	Image Analysis	3.10.

US Volume Blood Flow (VBF) Biomarker Committee:

A common testing protocol has been created and adapted to each round robin participating platform. The testing protocol has been and will continue to be utilized at multiple sites. These protocols and the same or similar phantoms will probably be employed in user and vendor compliance tests.

OBJECTIVE 6. COLLECT IMAGES AND ASSOCIATED DATA FOR A QIB DATA WAREHOUSE OR OTHER PUBLIC DATA REPOSITORIES, AND PERFORM ANALYSIS ON THE DATA TO SERVE QIB COMMITTEES AND THE BROADER IMAGING COMMUNITY.

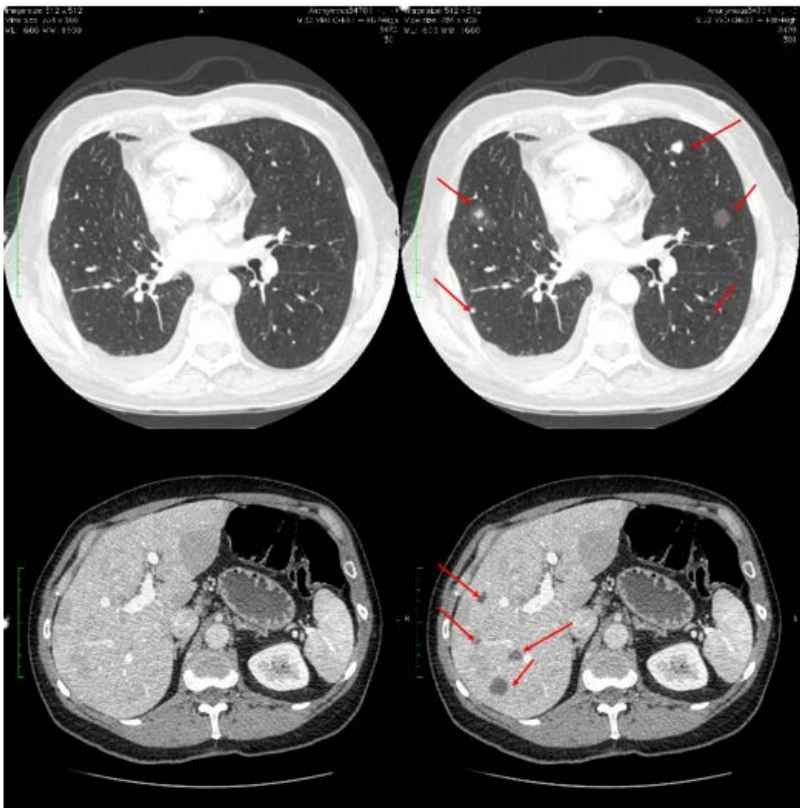
Note that, in general, data from all QIBA committee efforts have been, or will be, uploaded to the Quantitative Imaging Data Warehouse (QIDW) identified in the Methodology for Objective 6.

Selected specifics:

CT Modality:

Reference Image Set for Quantitation Conformance of Algorithmic Lesion Characterization, PI: Ehsan Samei, PhD - Duke University (CT Volumetry Biomarker Committee) - Round-5 (2015-16) Project C

A database composed of 100 patient thoracic images and 100 patient abdominal images was collected from de-identified cases at Duke University Hospital. Using the Duke database of lesion masks and their cloned renditions, 80 lesion morphologies (ground glass, part-solid, and solid with varying shapes and contrast and with sizes ranging within 4–20mm) were virtually inserted into each patient lung image set. All lesions contain realistic internal heterogeneity. Similar insertions were applied to liver images. Examples are shown below. Patient images and lesion models were examined by experienced radiologists. This database is packaged and will be uploaded to the QIDW.



Example reference hybrid lung (top row) and liver (bottom row) cases. Patient only images are on the left and patient images with inserted lesions are on the right. Lesions of various types, shapes, sizes, and contrasts were inserted in patient images.

A prior reference image set (*Reference Image Set for Quantification Conformance of Algorithmic Lesion Characterization*) without internal lesion heterogeneity has already been formed and its associated metadata are expected to be uploaded into QIDW in the coming months.

Reference Image Set for Lung Density Analysis Software Challenge, PI's: Charles Hatt, PhD (Imbio LLC) and Miranda Kirby, PhD (University of British Columbia), CT Lung Density Biomarker Committee – Reference human subject data from the COPDGene study for testing density analysis software performance and

QIBA Final Report - March 2018

variability. The data set will include 20 cases, representing a range of COPD severity (GOLD 1-4) and including both low and conventional dose CT scans.

MR Modality:

fMRI Biomarker Committee:

Quantitating Clinical fMRI Mapping of Language: Center, Spatial Extent, and Relative Strength of Active Areas, PI: James Voyvodic, PhD - Duke University, Jay. J. Pillai, MD - Johns Hopkins University (fMRI Biomarker Committee) - Round-5 (2015-16) Projects D1 and D2

This Round-5 language fMRI project allowed the upload of representative human fMRI data sets (in de-identified form) to the QIDW to support our reproducibility findings. We included examples of two language tasks from different subjects, representing different quantitative levels of reproducibility metrics.

PDF-MRI Biomarker Committee:

DWI-DRO Development for ADC Analyses, PI: Dariya Malyarenko, PhD - University of Michigan (PDF-MRI Biomarker Committee) – Round-5 (2015-16) Project G

Initial compatibility of the DWI-DRO was performed by University of Michigan investigators in application of select DICOM readers and analysis packages (GE FuncTool, Philips, QIBA_Phan, Osirix, IDL, and MatLab). Currently, the DWI-DRO and User Manual reside on the QIDW, though greater use and visibility may be achieved via links on the QIBA-Wiki.

Multi-site/vendor scans of QIBA DWI phantom and DWI Phantom Analysis Software (QIBA_Phan), supported in prior year groundwork projects, also reside on the QIBA QIDW and serve as a resources for the broader quantitative imaging community.

Ultrasound Modality:

US SWS Biomarker Committee:

Analysis of Sources of US SWS Measurement Inter-System Variability, PI: Mark Palmeri, MD, PhD – Duke University (SWS Ultrasound Biomarker Committee) – Round-5 (2015-16) Projects F1 and F2

Elastic and viscoelastic digital phantom data (finite element simulation data) previously uploaded to QIDW continue to be downloaded by academic and industry members of the QIBA community. Permissions for uploading SWS images and data from the Round-6 Project BB are being obtained from the two research sites. Obtaining permission for uploads of this nature is becoming extremely difficult due to increasing patient privacy concerns, but this time-consuming effort is continuing and the outcome is uncertain.

Simulation data sets have been developed and posted on the QIDW for use by research groups and manufacturers. The goal is to find approaches that allow different ultrasound systems to achieve the same SWS results from data generated using appropriate simulated visco-elastic materials. Simulated data representing elastic (lossless) and viscoelastic (tissue-mimicking) media have been released for download by interested parties, and several manufacturers have begun to look at the materials to determine if it is technically and economically feasible to analyze test data using their proprietary software. A comprehensive comparison of simulation results obtained with two common commercial finite element modeling software packages has been performed, and the corresponding code to process the data are available on GitHub. If this plan is successful, use of Digital Reference Objects (DROs) to analyze ways to achieve better agreement in SWS values will become possible.

Also, software has been developed for standardized reconstruction/processing of publicly archived shear wave speed data that will be verified and uploaded when testing is complete. This software will form the core of a reference standard for US SWS processing/measurement to be included in the next version of the Profile.

US Volume Blood Flow (VBF) Biomarker Committee:

Examination of Flow Phantom as Reference Standard for Validation of Ultrasound Volume Blood Flow Measurement, PI: Oliver Kripfgans, PhD - University of Michigan - Round-6 (2016-18 (NCE)) Project W

QIBA Final Report - March 2018

To date more than 18,450 volume datasets (500 GB) have been collected. These will be examined for compatibility for storage in the Quantitative Imaging Data Warehouse (QIDW).

OBJECTIVE 7. PROVIDE SUPPORT FOR THE QIB COLLABORATION PROGRAM STAFF (OUTSIDE ORGANIZATION STAKEHOLDERS AND QIB COALITION MEMBERS), PROJECT MANAGEMENT, MEETINGS, TRAVEL, AND CONFERENCE CALLS.

Support for all the above committee work, funded project management meetings, conference calls and travel continue to be administered and provided by the RSNA/QIBA staff, QIBA Chair / Vice-chair, and Scientific Liaisons. (Much of the administrative support for the US Volume Flow Biomarker Committee is provided by the American Institute of Ultrasound in Medicine in cooperation with the RSNA, and logistic support for the joint QIBA-European Imaging Biomarkers Alliance (EIBALL) MR Arterial Spin Labeling Task Force is being provided by the European Institute for Biomedical Imaging Research (EIBIR).)

B. COMBINED LIST OF GROUNDWORK PROJECTS APPROVED BY STEERING COMMITTEE ACROSS THE OBJECTIVES FOR ROUND-5 AND 6 (2015-2017) FUNDING.

Project Number	Biomarker Cmte	Round-5 & 6 Project Title	Investigator
A	PET Amyloid	Analyses to Support Amyloid Imaging Profile Development	Dawn Matthews, ADM Diagnostics, LLC
B	FDG-PET	Biologic and Reader Repeatability of FDG and CT Volumetric Parameters (ACRIN 6678 & MERCK)	Rathan Subramaniam, MD, PhD, MPH, Johns Hopkins University
C	CT Volumetry	Reference Image Set for Quantitation Conformance of Algorithmic Lesion Characterization	Ehsan Samei, PhD, Duke University
D1, D2	fMRI	Quantitating Clinical fMRI Mapping of Language: Center, Spatial Extent, and Relative Strength of Active Areas	James Voyvodic, PhD, Duke University Jay J. Pillai, MD, Johns Hopkins Univ
E	PET Amyloid	Amyloid Brain PET Test-Retest Meta-Analysis	Rathan Subramaniam, MD, PhD, MPH, Johns Hopkins University
F1, F2	SWS US	Analysis of Sources of US SWS Measurement Inter-System Variability	Mark Palmeri, MD, PhD, Duke Univ Shigao Chen, PhD, Mayo Clinic
G	PDF-MRI (DWI TF)	DWI-DRO Development for ADC Analysis	Dariya Malyarenko, PhD, University of Michigan
H	PDF-MRI	Dynamic Susceptibility Contrast MRI Phantom	Ona Wu, PhD, Harvard/Mass General Hospital
I	FDG-PET	A PET Metabolic Tumor Volume Digital Reference Object (PET-MTV-DRO)	Paul Kinahan, PhD, University of Washington
J	FDG-PET	A Procedure to Facilitate Greater Standardization of PET Spatial Resolution	Martin Lodge, PhD, Johns Hopkins University
K	Cross Modality	Aggregated Measures of Agreement for QIB Validation: An Open Source Toolkit	Daniel Barboriak, MD, Duke University
L	Lung Density	Investigation of Methods of Volume Correction for Lung Density CT	Sean Fain, PhD, University of Wisconsin-Madison

QIBA Final Report - March 2018

Project Number	Biomarker Cmte	Round-5 & 6 Project Title	Investigator
R1, R2	FDG-PET	SUV Quantification with Point Spread Function PET Reconstruction	Martin Lodge, PhD, Johns Hopkins University / Ronald Boellaard, PhD, University of Groningen
S	CT Vol	Methodology and Reference Image Set for Lesion Characterization in Terms of Texture and Morphology	Ehsan Samei, PhD, Duke University
T	PDF-MRI (DTI)	Measurements of Reproducibility of DTI Metrics on Clinical MR Scanners using a DTI Phantom	James Provenzale, MD, Duke University
U	Lung Density	CT Lung Density Biomarker: Translating Phantom Harmonization to Clinical Practice	Stephen Humphries, PhD, National Jewish Health
V1, V2	SPECT	Multi-center Phantom Study to Characterize Bias and Precision of Quantitative ¹²³I SPECT	Yuni Dewaraja, PhD, University of Michigan / John Dickson, PhD, University College – London Hospital
W	VBF	Examination of Flow Phantom as Reference Standard for Validation of Ultrasound Volume Blood Flow Measurement	Oliver Kripfgans, PhD, University of Michigan
X	SPECT	¹²³I DAT Scan Digital Reference Object Development	Robert Miyaoka, PhD, University of Washington
Y	PDF-MRI (DSC)	A Web-based Tool for Creating DSC Digital Reference Objects	Bradley Erickson, MD, PhD, Mayo Clinic
Z	PDF-MRI (DCE)	Evaluation of RF Transmit Calibration Options for Quantitative DCE-MRI	Krishna Nayak, PhD, University of Southern California
AA	PET Amyloid	Matched Digital and Physical Amyloid Phantom for Software and Scanner Validation	Paul Kinahan, PhD, University of Washington
BB	SWS	Establishing Acceptable Variance Limits for Healthy, F1 and ≥F2 Fibrosis Shear Wave Speed Values Across Systems and Between Operators for the QIBA Profile	Manish Dhyani, MD, Mass General Hospital
CC	FDG-PET	Simple Variability Estimates in PET	Timothy Turkington, PhD, Duke University
DD	PET Amyloid	Quantification of Reconstruction Method Impact on Measured Amyloid Load	Dawn Matthews, ADM Diagnostics, LLC

C. GENERAL PROGRESS ON ACTIVITIES BEYOND THE FUNDED PROJECTS

Additional updates from the committees are as follows.

CT Volumetry Biomarker Committee

The committee formed a group to explore the standardization of analysis of morphological features beyond volume, such as texture and shape, as biomarkers.

QIBA Final Report - March 2018

A CT liver phantom has been designed and fabricated. Image data were acquired from two different vendor 64-detector scanners per the designed imaging protocols. In total, we collected 384 image series (*i.e.*, 2 scanners x 96 imaging settings x 2 repeats) and 34.4 GB of image data. All of the acquired image data were transferred to the FDA for analysis and are ready to be submitted to the Quantitative Imaging Data Warehouse (QIDW). This section briefly outlines the statistical analysis used to assess the volume estimation performance of the segmentation and matched-filter algorithms. The statistical assessment was based on the metrology recommendations outlined by the QIBA Metrology Working Group, which included the analysis of bias, variance, reproducibility, and repeatability. The results showed that the Vendor A scanner had smaller bias compared to the Vendor B scanner. Both scanners showed similar RMSE levels, suggesting that the added bias associated with the Vendor B scanner was offset by having a smaller variance (stdev) compared to the Vendor A scanner. The reproducibility and repeatability results from the Vendor A scanner are similar to the Vendor B scanner except when mixed-density lesions are included. This is because the estimates for mixed-density lesions from the images obtained by Vendor A were not as accurate, compared to the rest of the data. This project is complete and a project report was given to QIBA. In addition, a manuscript titled *Volumetry of low-contrast liver lesions with CT: Investigation of estimation uncertainties in a phantom study* has been published in *Medical Physics* describing the study and providing more detailed results and discussions.

Further research into CT liver lesion volumetry was carried out to explore the effects of iterative reconstruction methods on measurement reliability. Using a heterogeneous (salt and gelatin) liver phantom with 10 spherical lesions (size from 5 to 20 mm in diameter, ~40HU contrast), results showed that accuracy and precision of lesions under 10 mm in diameter were affected by reconstruction technique: a model-based iterative reconstruction method had increased bias and reduced variability in comparison to FBP and statistical methods (which performed similarly). No statistically significant difference was observed for larger lesions. This work was presented at *SPIE Medical Imaging* in 2017, and is reported in the *2017 SPIE Medical Imaging Conference Proceedings* in an abstract titled *The effects of iterative reconstruction in CT on low-contrast liver lesion volumetry: a phantom study*.

The Lung Nodule Assessment in CT Screening Task Force has been working to ensure that the small nodule claims are consistent with the established claims of the advanced disease CT Volumetry Profile. Published results and unpublished data from members of the group have been used to inform development of Claim details. The committee has collaborated with manufacturer representatives to obtain technical parameter guidance for individual scanner models for quantitative applications. Public comments have been received, addressed, and the Profile was promoted to the Consensus stage in 4Q2017.

QIBA 3A Challenge: A new lesion quantitation challenge was designed and initiated by the QIBA CT 3A Group. The main objective of this challenge is to quantitatively benchmark the volume estimation performance of segmentation tools for lung lesions using phantom and clinical cases containing real and virtually-inserted lesions. The aim is to provide a quantitative understanding of the differences between segmentation approaches based on a reference data set, and further establish the statistical exchangeability between real and simulated lesions. Participants are asked to perform image-based segmentation on datasets generated using (1) an anthropomorphic phantom with synthetic and virtually inserted nodules and (2) clinical images containing real lung lesions and virtually inserted lesion models. Nodules were virtually inserted using three insertion methods: Techniques A, B, and C, where Technique A is a projection-domain insertion method, and Techniques B and C are image-domain insertion methods. Each participant is to use their segmentation algorithm to perform volume estimation on nodules with locations that are a priori provided. This Challenge remained open until March 1, 2017. Once completed, the results will be analyzed and presented at RSNA and prepared for publication.

CT Lung Density Biomarker Committee

Draft Profile and Claim developments are in progress, based on critical evaluation of literature for a lung density protocol. The acquisition and reconstruction specifications of CT images have been completed and are being evaluated by a working group of vendor scientists who are developing conformance procedures using the COPDGene Phantom. The image analysis section of the Profile is completed, and a software comparison is underway to establish repeatability of academic and commercial software packages that conform to the Profile.

QIBA Final Report - March 2018

The group has developed recommendations for reduced dose pulmonary quantitative CT (qCT) protocols to be used on multiple vendor scanners using automatic exposure control (AEC) and iterative reconstruction (IR). These protocols are guiding efforts to lower CT dose for ongoing and future clinical research studies and are currently being tested for feasibility in the ongoing multi-center longitudinal COPDGene Study (Project U).

A Task Force of CT vendor scientists has been formed to develop a conformance checklist and to suggest changes to the acquisition and reconstruction parameter specifications in efforts to mitigate measurement bias. The Task Force has organized a project that involves scanning the same COPDGene Phantom using three radiation doses on two models of each vendor's CT scanners. The CT Vendor Task Force has completed two rounds of scanning and is using these data to further improve harmonization and establish a joint conformance statement.

The Biomarker Committee has completed a meta-analysis of the CT lung density repeatability literature, thus finalizing their measurement repeatability claim for assessing emphysema progression. The meta-analysis may be the basis for a submission of a manuscript for publication in the peer-reviewed literature.

PDF-MRI Biomarker Committee

The majority of PDF-MRI activities are conducted within its respective Task Forces. Task Forces exist within the PDF-MRI Biomarker Committee for DCE-, DWI-, DTI-, and DSC-MRI. In addition, an ASL Profile is being developed in collaboration the European Society of Radiology's European Imaging Biomarkers Alliance (EIBALL) with administrative support from European Institute for Biomedical Imaging Research (EIBIR).

The DCE-MRI Task Force is presently focused on defining systematic literature review procedures, and their application for select organ/tumor sites, to support the DCE-MRI v2.0 Profile claims. Even within a given organ site, the literature often reveals a broad range of key acquisition parameters (e.g., temporal sampling rate) with incomplete description of methodology. Relatively uniform multi-site/-platform DCE-MRI methodology was achieved in the ACRIN 6701 test/retest prostate clinical trial used to field-test the DCE v1.0 Profile. Moreover, an automated software analysis package was developed and applied to data acquired on the QIBA DCE-MRI Phantom for site qualification in the ACRIN 6701 study. The analysis software and user manual have been uploaded to the QIDW, along with example data from scanners from three major MR system vendors and the associated reports produced by the software. In addition, an open-source software package to facilitate comparison of parametric images generated by different DCE-MRI analysis packages when utilizing the DRO, created as part of a previously funded groundwork project, is also available on the QIDW. This software is capable of importing 2D and 3D DICOM images, or binary data formats, as well as imaging formats such as TIFF and PNG. It generates difference and ratio maps (exportable as PNG), scatter diagrams and box-plots, and ANOVA statistics to more easily compare analysis packages. These resources are available on the QIDW for future clinical trials and the broader quantitative imaging community.

The DWI Task Force successfully completed scans of its physical QIBA DWI phantom, and all datasets were analyzed using the "QIBA_Phan" software developed as a prior groundwork project and available on the QIDW. In 2016, DWI Task Force members also completed development of a DWI DRO. Analogous to the DWI Task Force, the DSC Task Force is currently developing a physical phantom, protocol, and corresponding analysis software supported under QIBA groundwork project contracts. DTI Task Force leaders have previously developed a novel isotropic plus anisotropic physical diffusion phantom. This phantom allows ground-truth measurements of key DTI metrics: mean diffusivity (MD), fractional anisotropy (FA), radial diffusivity (RD) and axial diffusivity (AD). DTI Task Force members will continue scanning this phantom in the coming year to evaluate intra-/inter-platform variance of these metrics.

fMRI Biomarker Committee

The fMRI Biomarker Committee has completed its v1.0 sensorimotor pre-surgical mapping Profile (currently in Public Comment phase) and has moved on to pre-surgical mapping of eloquent brain tissue. Refinements to the clinical Claims and context were made, particularly the acquisition guidelines, as well as accompanying appendices with detailed performance specifications. Members are in the process of completing revisions to Section 3, Profile Activities, specific to the mapping of motor cortex. To inform conformance procedures, members are conducting groundwork studies focused on software analysis specifications.

QIBA Final Report - March 2018

The fMRI Bias Task Force meets bi-weekly to focus on the issue of bias in the fMRI measurement, center of mass of activation. This activity informs the Profile Claims definition and guides development of methodological sequences for image analysis that best achieve the claims.

fMRI-DRO testing was completed at 8 sites, all analyzing the same bilateral hand motion and language mapping DROs but with each site employing its own standard fMRI processing and analysis workflow. The activation map results accompanied by data analysis forms describing workflow were collected from each site. Generation and testing of advanced DROs for head motion in fMRI were performed. These include DROs from various combinations of selected empirical and synthetic datasets wherein amplitude and spatial distribution of task-related fMRI signals and associated fMRI noise were controlled. By fully specifying “ground truth” in this way, subsequent post-processing and display methods can be tested for the ability to accurately recover the original signal distributions and to quantify any inaccuracies that might be present. These DROs, containing realistic task signal and noise variability, including motion, performance, and neurovascular uncoupling (NVU) sources of variance, have been uploaded to the QIDW. These can be used for conformance testing and comparison of fMRI analysis and correction methods for coping with the variance of the BOLD signal in the primary motor cortex as a function of presence or absence of NVU.

Members of the fMRI Biomarker Committee contribute to the DICOM Working Group 16 fMRI Task Force. The proposed DICOM work item will build on recent quantitative imaging support added to the standard, with new elements created as necessary to represent fMRI acquisition, activation maps, and task paradigms. The functional requirements incorporated by WG-16 fMRI TF were drawn from work done in the QIBA fMRI Biomarker Committee.

MR Elastography (MRE) Biomarker Committee:

This BC did not require a funded groundwork project to collate adequate material to develop their hepatic MRE Profile, which is currently in Public Comment phase.

Proton Density Fat Fraction (PDFF) Biomarker Committee:

This BC did not require a funded groundwork project to collate materials for ongoing development of their hepatic PDFF Profile. A manuscript addressing the meta-analysis of PDFF measurements as compared to MR spectroscopy measurements to establish limits on bias, linearity, repeatability, and reproducibility has been published in *Radiology*.

FDG-PET/CT Biomarker Committee:

With the Technically Confirmed status of the FDG-PET/CT Profile completed, the FDG-PET/CT Biomarker Committee is now looking at two main topics: (1) Revisions to the Profile to include technical advancements, such as the use of physics modeling of the PET scanner in the image reconstruction process, and (2) steps to advance to the next stage of Clinically Confirmed. For the Clinically Confirmed stage, the Profile details will have been implemented in more than one facility and each participating system and person successfully met the specifications. In addition, the overall performance was determined and the Claim was achieved.

US SWS Biomarker Committee:

The goal of the SWS US Biomarker Committee is to develop a QIBA Profile for a single biomarker: ultrasound shear wave speed (SWS) as a measure of liver stiffness, which correlates with the degree of liver fibrosis/cirrhosis present. Major efforts center on completing groundwork studies and publishing results, continuing to understand and account for sources of bias in SWS estimation with ultrasound imaging systems, continuing to determine sources of variance in these estimates, minimizing those contributions, and finalizing the draft protocol and Profile documents.

Three recent areas of groundwork effort were: (1) validation of simulations and phantoms mimicking elastic and viscoelastic properties of liver, (2) comparison of SWS measurements in uniform liver-mimicking phantoms using ultrasound imaging systems, the established US non-imaging system, and, initially, MR elastography, and (3) assessing sources of measurement variability in shear-wave elasticity techniques. It is anticipated that the physical and digital phantoms will be part of these efforts, as well as of the conformance procedures.

QIBA Final Report - March 2018

Data have been acquired to investigate sources of variance from comorbidity, biological variability, and measurement methods that might affect SWS estimate correlation with fibrosis. Studies were based on a literature analysis of 1,548 publications, from which 102 SWS papers included a study of one or more confounding factors. A further analysis of the potential for steatosis and/or inflammation to affect the correlation of SWS with liver fibrosis was performed using results obtained for 242 subjects.

A standardized plan for archiving clinical and phantom data into the QIDW is being devised and will be included as an appendix to the QIBA Profile.

A first draft of the QIBA ultrasound Profile “SWS Estimation of Liver Fibrosis” has been created. Distribution within and approval by the Biomarker Committee is pending while the Profile document is converted to the new document template provided by the Process Committee and system-dependent methods descriptions are provided by the participating manufacturers. A standardized SWS data collection report form has been developed for inclusion in the Profile appendices.

US Volume Blood Flow (VBF) Biomarker Committee:

Beyond the initially three proposed manufacturers, two additional manufacturers were included in the project and are developing software on their scanners to facilitate the export of color flow and color power data in parallel, as well as the required geometry information needed for our computation. Furthermore, an additional platform from an already participating manufacturer was also integrated in the process for volume flow estimation.

With resources provided by the American Association of Physicists in Medicine (AAPM), we have purchased an additional phantom from Gammex Inc. to supplement the round robin study. In collaboration with Gammex Inc., we are examining a design for an MRI-compatible phantom for future MRI/US comparative testing.

CEUS Biomarker Committee:

A new Contrast Enhanced Ultrasound (CEUS) Biomarker Committee was initiated and progress is being made in establishing the extent of inter-system and inter-institution variability in flow measurements in phantoms. A biomarker committee is being formulated to develop an international consensus on a fetal maturity biomarker set.