

HHSN268201500021C, Quantitative Imaging Biomarkers Alliance (QIBA)

PROPOSED WORKPLAN

Consistent with the Proposed Methodology approved by the NIBIB Project Officer, QIBA has completed an initial project planning exercise covering the scope of work in the NIBIB award. The plan was developed based on a two-level review, with the first level comprising work by the several Biomarker Committees and Task Forces to define profiling activities and groundwork projects for submission to the Steering Committee, and the second level being a formal review, ranking, and approval by the Steering Committee of those projects deemed most likely to provide results against the contract objectives. This workplan is arranged in the following subsections:

- A. Review of activities responsive to each objective.
- B. Combined list of groundwork projects approved by steering committee across the objectives.
- C. For each modality-specific Coordinating Committee and Biomarker Committee, summary narratives for the projects reviewed, ranked and approved by the Steering Committee.

In the text that follows, Rounds 1 through 4 refer to projects funded from previous NIBIB contracts. Rounds 5 and 6 refer to activities that will be funded from the current contract. Round 5 refers to contract year 2015-16 and Round 6 refers to contract year 2016-17.

In the following sections, specific Round-5 (2015-16) groundwork projects are provided, where relevant, to highlight specific efforts in support of each objective. Additional groundwork projects will be funded in Round 6 (2016-17), and the development and selection of those additional groundwork projects will be accomplished using the same two-level process described above.

A. REVIEW OF ACTIVITIES RESPONSIVE TO EACH OBJECTIVE.

Those projects associated with the objectives are identified below. It should be noted that many of the projects are responsive to multiple objectives. In some cases, a given project is listed twice, but, in general, projects are referenced in the objective where they have the strongest contribution.

Note that, in general, the data from each of these efforts is available for uploading to the RSNA QIBA Quantitative Imaging Data Warehouse (QIDW), and that the committee work and funded projects are supported by the QIBA staff, Program Director, Principle Investigator/QIBA Chair, QIBA Vice Chair, and Scientific and Statistical Consultants.

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

QIBA Plans from Original Proposal: At least two Protocols or Profiles will be completed per contract year.

The work of the 11 QIBA Biomarker Committees and 11 Task Forces follows a defined, coordinated process described below to develop solutions and promote their adoption. QIBA efforts have produced three publicly-reviewed profiles to date: CT Volumetry for Advanced Disease, DCE-MRI for Solid Tumors, and FDG-PET for whole body studies.

Profiles from Biomarker Committees and Task Forces that are expected in 2016 or 2017 are noted below, and recently initiated Biomarker Committees that will become active during the funding period are also highlighted.

CT Coordinating Committee Efforts:

CT Volumetry Biomarker Committee: Starting from the released profile for use of CT volumetry as a response marker for drug development in advanced disease, a working group has been formed to develop a Profile for assessment of small nodules in the context of lung cancer screening. They are developing document text to define evidence-based consensus standards and processes for CT imaging in the setting of lung cancer screening, to allow for quantification of biologically meaningful longitudinal volume changes, with acceptable range of variance across vendor platforms. This activity is expected to result in the release a Profile in 2016.

QIBA Proposed Workplan

CT Lung Density Biomarker Committee: In parallel to the efforts described above, a Profile for the measurement of lung density based on quantitative CT measurements is in development and is expected to be released for public comment in late 2016 or early 2017.

CT Volumetry Biomarker Committee: Groundwork projects focused on volumetry of liver masses have been undertaken for the purposes of extending the first version of the CT Volumetry Profile to include quantitative assessment of liver masses. Phantoms developed in Round-3 and Round-4 groundwork projects, *i.e.*, funded by the previous NIBIB contract, were critical to this effort. It is anticipated that a profile for CT Volumetry of liver masses will continue to develop throughout the first year of the current contract, with a draft for public review in the latter part of the second year of funding.

MR Coordinating Committee Efforts:

fMRI Biomarker Committee: The QIBA BOLD fMRI Profile v1.0 has been developed to provide a systematic approach for optimizing Blood Oxygen Level Dependent (BOLD) fMRI brain mapping for treatment planning prior to surgery or invasive treatment intervention. Whereas the primary purpose of this Profile development is for individual patient care, application of the best practice guidelines it creates also has application to clinical trials. This Profile makes claims about the precision with which hemodynamic response in eloquent cortex, specifically motor cortex for the initial Profile, can be measured and displayed under a set of defined image acquisition, processing, and analysis conditions. It is anticipated this Profile should be released for public comment in 2016. The Biomarker Committee will then address language mapping, and a Round-5 (2015-16) groundwork project is focused on filling some of the gaps in knowledge needed in this area before a profile for such mapping can be drafted.

Perfusion/Diffusion/Flow-MRI Biomarker Committee: The PDF-MRI Biomarker has multiple Task Forces all focused on the development or revision of Profiles. The DCE-MRI Task Force is drafting Dynamic Contrast Enhanced (DCE)-MRI Profile v2.0, which will extend DCE-MRI Profile v1.0 by addressing 3.0T field strength-specific issues as well as parallel imaging acquisition issues affecting quantitative imaging. This effort builds on the success of Round-3 and Round-4 groundwork projects funded by the previous NIBIB contract. The goal is to have a version of the revised Profile available for public comment in 2016.

The DWI Task Force Profile v1.0 is being drafted to address technical performance standards for the acquisition of apparent diffusion coefficient (ADC) measurements in multiple organ systems. This effort builds on the success of Round-3 groundwork projects (previous NIBIB contract). The goal is to have a version of the Profile released for public comment by mid-2016.

Two newer Task Forces of the PDF-MRI Biomarker Committee are in earlier stages of development, including the Diffusion Tensor Imaging (DTI) Task Force and the Dynamic Susceptibility Contrast (DSC) MRI Task Force. These groups are seeking to have drafts of their respective profiles available for public comment in 2017.

MR Elastography Biomarker Committee: The MRE Biomarker Committee was formed in 2014 and is actively developing a draft Profile for quantitative MRE measurements in liver. As MRE measurements from all three major MR system vendors use very similar acquisition and post-acquisition data processing strategies, substantive groundwork has already been performed and results previously published. Therefore, it is anticipated the Profile development will progress rather rapidly and a draft Profile will be available for public comment in 2016.

Proton Density Fat Fraction MR Biomarker Committee: This committee was approved by the Steering Committee in November 2015 and is beginning its efforts, as with most Biomarker Committees, with a detailed review of the existing quantitative imaging biomarker literature. It is anticipated it will take approximately one to two years before this Biomarker Committee has a Profile ready to be released for public comment.

Nuclear Medicine Coordinating Committee Efforts:

FDG-PET Biomarker Committee: The *FDG-PET/CT as an Imaging Biomarker Measuring Response to Cancer Therapy* Profile is in the Publicly Reviewed status (December 2013). The next phase for the FDG-PET/CT Profile was field-testing, which was initiated at multiple sites and has led to suggested changes to the publicly reviewed Profile, as well as a revised checklist of procedure steps, and the FDG-PET/CT UPICT Protocol

QIBA Proposed Workplan

release in July 2014. It is anticipated that these changes will lead to a revision of the FDG-PET/CT Profile, which is expected to be ready for public review in early 2017.

PET-Amyloid Biomarker Committee: The second PET Profile in development is for amyloid plaque neuroimaging. It is anticipated this Profile will utilize the framework of the FDG-PET/CT Profile and the recent efforts of the amyloid plaque neuroimaging community development process, and will lead to a version to be released for public comment in 2016 or early 2017. These efforts will be greatly aided by the two Round-5 (current NIBIB contract) groundwork projects funded in 2015-16 (see Sections B and C).

SPECT Biomarker Committee: The SPECT Biomarker Committee was approved by the Steering Committee in 2015 and is beginning its efforts, as with most Biomarker Committees, with a detailed review of the existing quantitative imaging biomarker literature. It is anticipated it will take approximately one to two years before this Biomarker Committee has a Profile ready to be released for public comment.

Ultrasound Coordinating Committee Efforts:

Shear Wave Speed Biomarker Committee: An initial draft of the Shear Wave Speed (SWS) Profile is in development. The literature review for the Profile has been completed, and approaches to reducing differences between the different SWS-measuring systems using new viscous phantoms and simulations have been characterized. Comparisons between liver fibrosis measurements using SWS and other techniques, including MRI-based and non-imaging based techniques, have been completed using previous NIBIB contract funds. An initial Profile for public review is targeted for early-to-mid-2016.

Volume Flow Ultrasound Biomarker Committee: This committee was approved by the Steering Committee in fall 2015 and is beginning its efforts, as with most Biomarker Committees, with a detailed review of the existing quantitative imaging biomarker literature. It is anticipated it will take approximately one to two years before this Biomarker Committee has a Profile ready to be released for public comment.

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

QIBA Plans from Original Proposal: At least two Profiles will be tested or revised per contract year.

CT Coordinating Committee:

The CT Volumetry Biomarker Committee is evaluating options for a formal field test of the publicly reviewed *CT Tumor Volume Change* Profile (v2.2). Such efforts include communications with cooperative trial groups, as was done with ACRIN for field testing of the *DCE-MRI Quantification* Profile (v1.0), and with imaging core labs.

MR Coordinating Committee:

The PDF-MRI Biomarker Committee DCE-MRI Task Force anticipates the receipt of the results of the field test of *DCE-MRI Quantification* Profile (v1.0) recommendations, which were implemented in ACRIN 6701 *Repeatability Assessment of Quantitative DCE-MRI and DWI: A Multicenter Study of Functional Imaging Standardization in the Prostate*. As of October 29, 2015, 21 of the target of 30 evaluable patients had been accrued from seven sites representing all three major MR vendors (GE, Philips, Siemens). The study is anticipated to conclude in mid 2016. In addition, based on groundwork projects in Round-3 and Round-4, *i.e.*, previous NIBIB contract funds, the Task Force is revising the v1.0 Profile to include 3.0T field strength and parallel imaging. It is anticipated the revised profile will be released for public comment in 2016.

The PDF-MRI Biomarker Committee DWI-MRI Task Force has completed its literature review and Round-3 groundwork projects (focused on development and dissemination of a NIST-traceable DWI MRI Phantom and associated data analysis software and funded by a previous NIBIB contract) and is preparing a DWI-MRI Profile that is expected to be released for public comment by mid-2016. Key aspects of the draft Profile have been implemented in collaborative studies with the EORTC / IMI and with a group of São Paulo neuroradiologists leading a multicenter, multivendor clinical trial of DWI in glioblastoma patients to allow initial field testing of Profile recommendations.

Nuclear Medicine Coordinating Committee:

QIBA Proposed Workplan

FDG-PET Biomarker Committee: The FDG-PET/CT Profile was field-tested at multiple sites and the results of the testing led to suggested changes to the publicly reviewed Profile, as well as a revised checklist of procedure steps. It is anticipated that a revision of the FDG-PET/CT Profile will be ready for public comment in early 2017

Ultrasound Coordinating Committee:

Field testing and/or revision of QIBA ultrasound Profiles is in an earlier stage of development due to the more recent initiation of the ultrasound SWS Biomarker Committee. The SWS Ultrasound Biomarker Committee anticipates release of the draft Profile for public comment in early-to-mid-2016.

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, NUCLEAR MEDICINE, AND ULTRASOUND.

QIBA Plans from Original Proposal: At least one candidate project from each modality, and at least six total projects that will address separate QIBA Profiles, will be funded and completed.

Initial bias and variance measures for each biomarker are extracted from the existing literature, and gaps in the data necessary to understand the sources of bias and variability are noted and lead to funded groundwork projects necessary to obtain data that may be used to fill such gaps. The following are groundwork projects proposed by the modality-specific Coordinating Committees (and one cross-modality project) and reviewed and prioritized by the Steering Committee for 2015-2016 (Round-5). The combined list of groundwork projects is provided in Section B, and narrative descriptions of the projects are provided in Section C.

CT Coordinating Committee:

Reference Image Set for Quantification Conformance of Algorithmic Lesion Characterization, PI: Ehsan Samei, PhD – Duke University (CT Volumetry Biomarker Committee)

Investigation of Methods of Volume Correction for Lung Density CT, PI: Sean Fain, PhD – University of Wisconsin (Lung Density Biomarker Committee)

MR Coordinating Committee:

DWI-DRO Development for ADC Analysis, PI: Dariya Malyarenko, PhD – University of Michigan (PDF-MRI Biomarker Committee / DWI Task Force)

Dynamic Susceptibility Contrast MRI Phantom, PI: Ona Wu, PhD – Harvard University / Massachusetts General Hospital (PDF-MRI Biomarker Committee / DSC-MRI Task Force)

Quantitating Clinical fMRI Mapping of Language: Center, Spatial Extent, and Relative Strength of Active Areas, PI: James Voyvodic – Duke University (fMRI Biomarker Committee)

Nuclear Medicine Coordinating Committee:

Analyses to Support Amyloid Imaging Profile Development, PI: Dawn Matthews – ADM Diagnostics, LLC (PET Amyloid Biomarker Committee)

Amyloid Brain PET Test-Retest Meta Analysis, PI: Rathan Subramaniam, MD, PhD, MPH – Johns Hopkins University (PET Amyloid Biomarker Committee)

A Procedure to Facilitate Greater Standardization of PET Spatial Resolution, PI: Martin Lodge, PhD – Johns Hopkins University (FDG-PET Biomarker Committee)

Biologic and Reader Repeatability of FDG and CT Volumetric Parameters (ACRIN 6678 & MERCK), PI: Rathan Subramaniam, MD, PhD, MPH – Johns Hopkins University (FDG-PET Biomarker Committee)

A PET Metabolic Tumor Volume Digital Reference Object (PET-MTV-DRO), PI: Paul Kinahan, PhD – University of Washington (FDG-PET Biomarker Committee)

QIBA Proposed Workplan

Ultrasound Coordinating Committee:

Analysis of Sources of US SWS Measurement Inter-System Variability, PI: Mark Palmeri, MD, PhD – Duke University (SWS Ultrasound Biomarker Committee)

Cross-Modality Groundwork Project:

A twelfth groundwork project, which is cross-modality in nature, was also recognized and prioritized by the Steering Committee for Round-5 funding:

Aggregated Measures of Agreement for QIB Validation: An Open Source Toolkit, PI: Daniel Barboriak, MD – Duke University

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 QIB PROFILES.

QIBA Plans from Original Proposal: At least two groundwork projects focused on physical and/or virtual (DRO) phantoms will be funded and completed.

The following Round-5 groundwork projects were funded to develop and/or utilize phantoms in support of Profiles being developed for public comment or for Profiles undergoing revision:

CT Coordinating Committee:

Reference Image Set for Quantification Conformance of Algorithmic Lesion Characterization, PI: Ehsan Samei, PhD – Duke University (CT Volumetry Biomarker Committee)

MR Coordinating Committee:

DWI-DRO Development for ADC Analysis, PI: Dariya Malyarenko, PhD – University of Michigan (PDF-MRI Biomarker Committee / DWI Task Force) – A synthetic (Digital Reference Object) phantom.

Dynamic Susceptibility Contrast MRI Phantom, PI: Ona Wu, PhD – Harvard University / Massachusetts General Hospital (PDF-MRI Biomarker Committee / DSC-MRI Task Force) – A physical phantom.

Nuclear Medicine Coordinating Committee:

A PET Metabolic Tumor Volume Digital Reference Object (PET-MTV-DRO), PI: Paul Kinahan, PhD – University of Washington (FDG-PET Biomarker Committee) – A synthetic (Digital Reference Object) phantom.

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

QIBA Plans from Original Proposal: A general set of procedures and processes to demonstrate conformance with QIBA Profiles will be developed, and at least two groundwork projects focused on conformance testing will be funded and completed.

A key focus for the various profiling efforts this year is to address issues of conformance, including the processes and procedures to demonstrate it. Many of these efforts build upon prior Round-3 and Round-4 groundwork projects with respect to the development of physical and synthetic (DRO) phantoms and associated data analysis software. For the current contract, some of the funded Round-5 groundwork projects specifically include development of additional predictive metrics for use in calibration and quality control programs and development of evaluation procedures to verify conformance by vendors and providers of service with QIBA Profiles. These Round-5 projects include the following:

CT Coordinating Committee:

Reference Image Set for Quantification Conformance of Algorithmic Lesion Characterization, PI: Ehsan Samei, PhD – Duke University (CT Volumetry Biomarker Committee)

QIBA Proposed Workplan

MR Coordinating Committee:

DWI-DRO Development for ADC Analysis, PI: Dariya Malyarenko, PhD – University of Michigan (PDF-MRI Biomarker Committee / DWI Task Force) – A synthetic (Digital Reference Object) phantom

Dynamic Susceptibility Contrast MRI Phantom, PI: Ona Wu, PhD – Harvard University / Massachusetts General Hospital (PDF-MRI Biomarker Committee / DSC-MRI Task Force) – A physical phantom

Nuclear Medicine Coordinating Committee:

A PET Metabolic Tumor Volume Digital Reference Object (PET-MTV-DRO), PI: Paul Kinahan, PhD – University of Washington (FDG-PET Biomarker Committee) – A synthetic (Digital Reference Object) phantom

Ultrasound Coordinating Committee:

Analysis of Sources of US SWS Measurement Inter-System Variability, PI: Mark Palmeri, MD, PhD – Duke University (SWS Ultrasound Biomarker Committee)

Cross-Modality Groundwork Project:

Aggregated Measures of Agreement for QIB Validation: An Open Source Toolkit, PI: Daniel Barboriak, MD – Duke University

It is anticipated that additional groundwork projects focused on conformance assessment will be initiated in 2016-17 (Round 6 groundwork projects).

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

QIBA Plans from Original Proposal: At least one groundwork project focused on providing data for the Quantitative Imaging Data Warehouse (QIDW) will be funded and completed.

At this time, the QIDW hosts digital reference objects (DROs) from groundwork projects completed in Rounds-1-4 (2013-2015) as well as data from collaborations between QIBA and the EORTC/IMI and NIST/ISMRM. During the 2015-16 period, additional datasets will be added to the QIDW from groundwork projects that began in prior rounds of funding as well as during the Round-5 (2015-16) period. Of the funded Round-5 groundwork projects, data will be provided to the QIDW from the following projects:

CT Coordinating Committee:

Reference Image Set for Quantification Conformance of Algorithmic Lesion Characterization, PI: Ehsan Samei, PhD – Duke University (CT Volumetry Biomarker Committee) – Reference image set data and associated metadata will be uploaded to the QIDW upon completion of groundwork project.

MR Coordinating Committee:

DWI-DRO Development for ADC Analysis, PI: Dariya Malyarenko, PhD – University of Michigan (PDF-MRI Biomarker Committee / DWI Task Force) – A synthetic (Digital Reference Object) phantom. This DRO data, like all other MR Coordinating Committee DROs, will be uploaded to the QIDW upon completion of groundwork project.

Dynamic Susceptibility Contrast MRI Phantom, PI: Ona Wu, PhD – Harvard University / Massachusetts General Hospital (PDF-MRI Biomarker Committee / DSC-MRI Task Force) – A physical phantom which will be scanned at multiple centers on platforms from multiple vendors with data and analyses subsequently uploaded to the QIDW upon completion of the project.

Quantitating Clinical fMRI Mapping of Language: Center, Spatial Extent, and Relative Strength of Active Areas, PI: James Voyvodic – Duke University (fMRI Biomarker Committee)

Nuclear Medicine Coordinating Committee:

QIBA Proposed Workplan

A PET Metabolic Tumor Volume Digital Reference Object (PET-MTV-DRO), PI: Paul Kinahan, PhD – University of Washington

Ultrasound Coordinating Committee:

Analysis of Sources of US SWS Measurement Inter-System Variability, PI: Mark Palmeri, MD, PhD – Duke University (SWS Ultrasound Biomarker Committee)

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

QIBA Plans from Original Proposal:

Project Management:

- Implement and oversee QIBA project-funding-request applications, review and selection processes.
- Create and maintain tracking process for all selected projects/subcontracts.
- Solicit and consolidate information/data from QIBA committees to support online, semi-annual and annual government reporting.

Conference Call support:

- Schedule calls, disseminate information, and prepare meeting materials and summaries for the teleconference meetings of all of the QIBA collaboration committees.

Communications

- Create and update collaboration web and wiki pages, rosters, and governance documents.
- Produce QIBA collaboration Newsletter.

In-Person Meeting Support

- Procure meeting space and negotiate contracts for services.
- Create registration mechanisms for all in-person meetings of the QIBA collaboration.
- Provide logistical support for speakers and attendees, including travel and hotel arrangements, and onsite support.
- Create and maintain meeting websites.

Quantitative Imaging Data Warehouse

- Coordinate and manage support and operation of the image archive.
- Implement and monitor application and review processes for upload and access to data.
- Coordinate and assist with curation of data.

Objective 7 includes the organizing of scientific meetings (Steering Committee and QIBA's Annual Meeting), and various Coordinating Committee, Biomarker Committee, and Task Force teleconferences at a wide range of frequencies (weekly, biweekly, monthly, and quarterly) depending on the specific nature of the groups. Additionally, it encourages the dissemination of QIBA documentations and Profiles broadly across the scientific, clinical, and industrial communities, by members of the QIBA coalition traveling to various scientific conferences to present QIBA results, both domestically and internationally. These activities are considered incidental to the requirement.

QIBA Proposed Workplan

B. COMBINED LIST OF GROUNDWORK PROJECTS APPROVED BY STEERING COMMITTEE ACROSS THE OBJECTIVES FOR ROUND-5 (2015-2016) FUNDING.

Project Number	Biomarker Cmte	Project Title	Investigator
A	PET Amyloid	Analyses to Support Amyloid Imaging Profile Development	Dawn Matthews, ADM Diagnostics, LLC
G	PDF-MRI (DWI TF)	DWI-DRO Development for ADC Analysis	Dariya Malyarenko, PhD, University of Michigan
F1, F2	SWS US	Analysis of Sources of US SWS Measurement Inter-System Variability	Mark Palmeri, MD, PhD, Duke University
H1, H2	PDF-MRI	Dynamic Susceptibility Contrast MRI Phantom	Ona Wu, PhD, Harvard/Mass General Hospital
K	Cross Modality	Aggregated Measures of Agreement for QIB Validation: An Open Source Toolkit	Daniel Barboriak, MD, Duke University
E	PET Amyloid	Amyloid Brain PET Test-Retest Meta Analysis	Rathan Subramaniam, MD, PhD, MPH, Johns Hopkins University
C	CT Volumetry	Reference Image Set for Quantification Conformance of Algorithmic Lesion Characterization	Ehsan Samei, PhD, Duke University
J	FDG-PET	A Procedure to Facilitate Greater Standardization of PET Spatial Resolution	Martin Lodge, PhD, Johns Hopkins University
D1, D2	fMRI	Quantitating Clinical fMRI Mapping of Language: Center, Spatial Extent, and Relative Strength of Active Areas	James Voyvodic, PhD, Duke University
B	FDG-PET	Biologic and Reader Repeatability of FDG and CT Volumetric Parameters (ACRIN 6678 & MERCK)	Rathan Subramaniam, MD, PhD, MPH, Johns Hopkins University
I	FDG-PET	A PET Metabolic Tumor Volume Digital Reference Object (PET-MTV-DRO)	Paul Kinahan, PhD, University of Washington
L	Lung Density	Investigation of Methods of Volume Correction for Lung Density CT	Sean Fain, PhD, University of Wisconsin

C. FOR EACH MODALITY-SPECIFIC COORDINATING COMMITTEE AND BIOMARKER COMMITTEE, SUMMARY NARRATIVES FOR THE PROJECTS REVIEWED, RANKED AND APPROVED BY THE STEERING COMMITTEE. (ROUND-5 GROUNDWORK PROJECTS, 2015-16)

CT COORDINATING COMMITTEE

CT Volumetry Biomarker Committee:

Reference Image Set for Quantification Conformance of Algorithmic Lesion Characterization, PI: Ehsan Samei, PhD – Duke University

Project Description

Our ability to accurately characterize a lesion is fundamentally limited by the varying confluence of anatomical background surrounding the lesion. As such, simplistic phantoms based on uniform or fixed constructs have limited relevance to clinical conditions in quantifying a varying lesion. Trials based on actual clinical images are in a sense ideal, but they also suffer from a lack of ground truth: it is nearly impossible to fully know the true size and morphology of a lesion in vivo, as subsequent pathological validations are rare, and even if done, the lesion can be deformed in the resection process.

The objective of the present project is to create and to make publicly available an image dataset of clinical CT scans with inserted synthetic, realistic lesions (Fig. 1). The dataset provides anatomical variability as exists in actual clinical datasets, while the same time the synthetic nature of the lesions offers the advantage of known truth. Multiple datasets can potentially be rendered as varying lesions can be reinserted countless times various locations, thus “refreshing” the test set. The lesions can be multiplied into additional clones with statistical variability, yet share the same generalized properties as those of the modeled lesions.

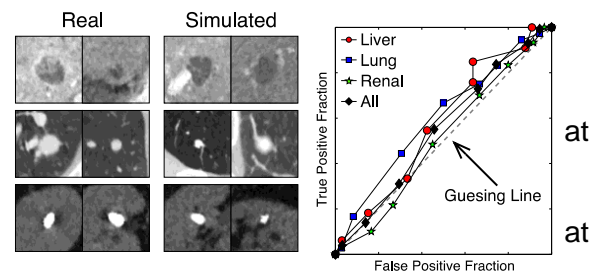


Fig. 1: Simulated 3D liver (top), lung (middle), and kidney stone (bottom) lesions are indistinguishable from real lesions based on expert observer studies (right).

The deliverables for the project are 1) a (static) reference database of 100 hybrid clinical image sets with lung nodules, 2) a (static) reference database of 100 hybrid clinical image sets with liver lesions, both data sets with confirmed and validated added lesions for gold-standard quantitative evaluation, and 3) a software tool for (dynamic) lesion addition and database creation upon demand.

Specific Aims

Aim 1. Establish statistical exchangeability of hybrid and real datasets for lesion quantification (mo. 1-6). In this aim, we first create an optimum hybrid database with realistic lesions based on methods developed in a Round 4 project. The lesions will be evaluated in terms of volume accuracy and precision. The precision will be compared with those from evaluation of “coffee break.” The comparison provides further evidence that the hybrid dataset can be used as a surrogate to clinical datasets (with the added availability of the complete knowledge of the lesion characteristics).

Aim 2. Create static reference clinical datasets with known truth (inserted realistic lesions) for validation of CT volumetry (mo. 3-10). With confirmation of Aim 1, in this aim, 100 normal images from LIDC will be inserted with 100+ lung lesions. A corresponding 100 normal abdominal CT scans will be inserted with 100+ liver lesions. For this aim, we will rely on Duke database of lesion masks and their cloned renditions. The lesion addition will take place in image projection domain, using the adjusted and validated techniques developed in our Round 4 project. In terms of realism, the simulated lesions have already undergone one clinical evaluation by experienced radiologists, indicating that they are undistinguishable from real lesions (Fig. 1), and statistical exchangeability test of Aim 1. However, for each case, the quality and location of the added lesions will be verified by experienced radiologists in terms of realism and suitability. This reference dataset will be used to conduct evaluation of quantitative performance across a select series of representative commercial and research software packages for nodule volumetry. The databases will be made publicly available so other groups will be able to benchmark their volumetry software using a validated reference clinical set without the need for additional image acquisition.

QIBA Proposed Workplan

Aim 3. Create a dynamic hybrid platform building towards establishing statistical thresholds of quantifications (mo. 9-12). In the final aim of this project, we will create a software tool for lesion addition and database creation upon demand. The software will use a normal clinical dataset and desired lesion attributes and locations (fixed or variable based on a *priori* statistical distributions) as input to create a dynamic dataset with dynamic characteristics as defined by the operator (Fig. 2). It is envisioned that while the static databases of Aim 2 offer conformance with pre-defined

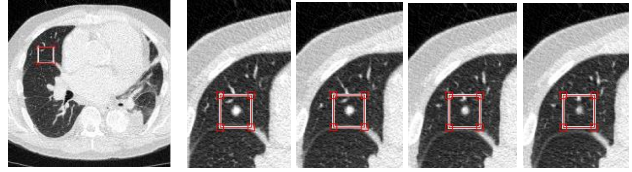


Fig. 2: A target image set (left) with an inserted lung nodule in consecutive slices (right).

performance thresholds, the software of Aim 3 can be used towards ascertaining operational thresholds above which specific algorithms provide statistically stable volumetric performance.

Lung Density Biomarker Committee:

Investigation of Methods of Volume Correction for Lung Density CT, PI: Sean Fain, PhD – University of Wisconsin

Project Description:

CT exams have shown utility in evaluating severity and progression of pulmonary emphysema in patients with chronic obstructive pulmonary disease (COPD). One of the major tasks in the profile drafting effort by the QIBA Lung Density Biomarker Committee is to assess the lung density bias and precision of repeat CT scans, using published longitudinal studies on non-diseased subjects. The repeatability coefficient determined from each study, a criterion for quantifying a real change in a lung density metric, has been shown to improve across all studies after volume adjustment to account for the change in the apparent lung density due to respiration. While there is consensus in the necessity of volume adjustment to achieve more accurate interpretation of densitometry CT measurements, there are competing methods and procedures among the published studies. Our attempts to evaluate the merit of a given method of volume adjustment were often hindered by the lack of details in the report. The prevailing models in use are of statistical in nature, making comparisons difficult without the raw data. Contacting authors of those studies had limited success. To understand the problem, we have performed some analysis with a set of the NLST data by applying statistical modeling at the sample level as most studies did which yielded similar improvement in the repeatability coefficient. However, the NLST data were taken one year apart, much longer the 3 month criterion for the studies included in the meta-analysis for the Profile document writing. A longer duration would likely produce real changes in the lung density due to natural progression. Moreover, they were taken on a 4-slice scanner which is outdated. We have further requested access to the COPDGene repeat-scan data sets with a range of intervals, which would allow us to create statistical and physiological models to make comprehensive evaluations of various methods of lung volume adjustment and its effect on the repeatability coefficient in particular and all sources of variances in general.

Primary goals and objectives:

Lung density metrics such as RA950 and Perc15 varies linearly with the change in total lung volume as determined by CT volume histograms. By correlating the change of a metric to the change of volume at the sample level, a statistical model can be constructed to realign the respiration levels at the baseline and repeat scans. However, the underlying assumption of conservation of lung mass (sponge model) may not be valid due to the blood flow in the vessels during respiration. Dr Philip Judy has investigated the hypothesis that the mass estimated from lung parenchyma partition of the does not vary during respiration and the histogram variation is associated the vessel-wall part of the histogram. By modeling the lung histograms the relationship between any given density metric and volume can be quantified with statistical parameters.

Deliverables:

QIBA Proposed Workplan

- 1) Quantify the validity of the sponge model based on the constant mass assumption by testing various statistical models for volume adjustment to assess the effects on the metrics above, using the existing COPDGene data set and new data acquired on a compressible lung phantom.
- 2) Quantify the invalidity of the constant mass assumption by partitioning lung parenchyma and vessel-wall in the histogram, and construct new models for volume adjustment, and reassess the effects on the metrics.
- 3) The results from 1) and 2) outcomes may result in a recommendation for the most favorable method of volume adjustment, a crucial step in making CT lung densitometry a quantitative tool for the assessment of COPD disease severity and progression.

MR COORDINATING COMMITTEE

PDF-MRI Biomarker Committee:

DWI-DRO Development for ADC Analysis, PI: Dariya Malyarenko, PhD – University of Michigan

Project Description:

Quantitative metrics provided by DWI, such as ADC, are being incorporated into clinical trials as potential radiological markers for diagnosis or therapy response. The sources of variability for such metrics include both DWI acquisition and post-processing. The quality control of DWI acquisition is being addressed by development of DWI phantoms and MRI equipment evaluation protocols. To assess and eliminate variability from post-processing, diverse DWI analysis SW and tissue models utilized by the clinical sites needs to be tested using uniform (vendor-agnostic) DICOM DWI standard Digital Reference Object (DRO). To provide the means to benchmark DWI SW performance across sites against ground truth over the tissue-relevant range of ADC values, this project proposes to build trace-DWI DRO using forward tissue diffusion models/values with Rician noise and standard diffusion DICOM attributes. The feedback from such DRO will facilitate optimization of DWI analysis and acquisition for specific tissue diffusion model and establishment of corresponding DWI DICOM requirements for vendors.

Primary goals and objectives:

Similar, to DCE DRO, the DWI DRO will be designed as set of trace-DWI DICOM images that cover the tissue-relevant range of input ADC values and SNR levels for single-component (mono-exponential) diffusion as a function of b-values (“dynamic” scans). The DRO DICOM header will utilize acquisition parameters used for QIBA DWI phantom assuming standard MRI acquisition bit depth of 12. Once built, the mono-exponential DRO framework will provide standard DW-MRI DICOM data attributes for the array of b-values used in clinical studies and could be expanded to multi-component diffusion models (e.g., bi-exponential or IVIM).

Deliverables:

1. Definition of parameter space suitable for DRO derivation
2. Definition of DICOM-compliant trace-DWI DRO attributes
3. DRO DICOM generation using defined diffusion models and input parameters, including noise
4. Test procedures to evaluate DRO by reproducing input parameters and models
5. Three-D DWI DRO standard DICOM set with analysis and performance evaluation instructions

Timeline [must include intermediate measureable milestones]:

DRO development will begin when funding has been awarded with initial proposal of DRO structure, sampled DWI parameter space and output DICOM format available for QIBA/PDF approval within 3 months. Creation of MatLab routines to generate DRO DICOM that meets design targets will require approximately 6 months. It is anticipated that initial DRO testing and DICOM compliance will be performed using open-source quantitative imaging community tools, like 3D Slicer. The remaining 3 months will be used to finalize DRO DICOM and corresponding testing procedures to evaluate performance for arbitrary DWI analysis SW.

QIBA Proposed Workplan

Dynamic Susceptibility Contrast MRI Phantom, PI: Ona Wu, PhD – Harvard University / Massachusetts General Hospital

Project Description:

Dynamic Susceptibility Contrast (DSC) MRI is routinely used in the clinics to image brain tumors. It has been hypothesized that changes in relative cerebral blood volume (rCBV) can be used as a biomarker to evaluate the efficacy of novel anti-tumor treatments. Differences in measured rCBV values over time may be due to both physiological and non-physiological factors, and therefore non-physiological variations need to be measured and limited. For multi-center clinical trials, stability of image acquisition across centers, vendors and time will be critical. We propose to design and to prototype a DSC phantom, and to perform phantom studies to evaluate the temporal and spatial stability of a DSC phantom using a “generic” acquisition protocol across multiple centers to assess reproducibility of these measurements. This information can then be used to provide estimates of bias and reproducibility that can be used to expedite the development of a DSC Profile. The phantom and protocols can ultimately be used for site qualification in multicenter clinical trials and for quality control in routine clinical practice.

This proposal is timely since DSC, and in particular rCBV, has already been shown to be a prognostic marker for overall survival in a recent trial of bevacizumab in GBM patients involving 21 patients and 5 sites. The results are encouraging, but lacking some confidence until an appropriate DSC phantom is available to assess reproducibility of these measurements. To make such studies feasible on a larger scale involving more sites while maintaining high quality, a DSC phantom will be important. In addition, if DSC is to be used routinely in the clinics either to triage patients or to monitor treatment effects of antiangiogenic drugs, a phantom such as we propose will be essential for quality control. We will work with the different vendors to create a generic acquisition protocol. To perform the multicenter reproducibility study, we will scan the two prototype phantoms at 5 centers at two time-points one-week apart.

Co-investigators will include Ona Wu PhD (Massachusetts General Hospital), Bradley Erickson MD PhD (Mayo Clinic), Matthias van Osch PhD (Leiden University Medical Center), John Kirsch PhD (Siemens), Karl Stupic PhD (NIST) and Kathryn Keenan PhD (NIST).

Primary goals and objectives:

Our primary goal is to develop a prototype DSC phantom from which a gradient of susceptibility values can be measured. We will work closely with NIST, building upon experience gained from existing NIST phantoms such as the DWI and DCE phantoms. NIST already has a Nano-Iron Phantom in development using iron oxide nanoagents in gels. Our secondary goal will be to create generic acquisition protocols by which we can assess the contrast-to-noise of the susceptibility measurements as well as stability across time across multiple vendors. Our third goal will be to estimate reproducibility and feasibility of performing these measurements across multiple centers (5 sites) at multiple time points (one-week apart). Prototype phantoms will be shipped to participating sites and data will be analyzed

Deliverables:

Our deliverables will consist of two DSC phantoms, generic acquisition protocols for Siemens, Philips and General Electric 1.5T and 3T MRI scanners, data analysis software and estimates of reproducibility of phantom measurements across 5 centers acquired at each center 1 week apart.

fMRI Biomarker Committee:

Quantitating Clinical fMRI Mapping of Language: Center, Spatial Extent, and Relative Strength of Active Areas, PI: James Voyvodic – Duke University

Project Description

Version 1.0 of the QIBA Profile on quantitative fMRI has focused on mapping motor cortex function. Because motor cortex mapping is relatively straightforward in terms of both anatomical organization and the scan protocols and tasks involved, focusing on motor mapping has allowed us to analyze many important sources of variance in fMRI in a simple brain system. Our Year-4 QIBA-funded project is using synthetic digital reference

QIBA Proposed Workplan

objects (DROs) to quantitate the influence of scanner noise, head motion, task performance, and neurovascular uncoupling on the reproducibility and bias of fMRI, using relatively simple models of brain activity as seen in the motor cortex. That effort will allow us to complete our first Profile, specifying quantitative constraints on scanner SNR, subject compliance (motion and performance), and tissue neurovascular coupling (NVC) necessary to achieve our Claim for quantitative localization of the center of hand motor cortex. Identifying such quantitative data quality constraints is an important advance in fMRI.

Having established this first claim for localizing centers of activations, the next step is to address the other two quantitative claims central to fMRI mapping: measuring the spatial extent of active cortex and the relative activation amplitude of different areas activated (e.g. hemispheric dominance). These questions are less important in motor mapping but they are critically important for mapping language areas of the brain. In Year-5, therefore, we propose to address these language mapping issues in order to add these two important quantitative claims to our Version 1.1 fMRI Profile. To do so we will use approximately 100 patient and healthy volunteer fMRI sessions in which each subject has performed a language mapping task more than once within the same scan session. We will apply the optimized analysis methods identified by our DRO studies to preprocess and identify active areas in each language scan. For each subject we will use the maps of statistically significant activation to extract BOLD amplitude and statistical parameters in 4 speech-related ROIs (posterior temporo-parietal cortex and inferior frontal cortex in both hemispheres). We will parameterize the spatial distribution of signal amplitude within each ROI using the AMPLE algorithm as in our Year-1 Reproducibility studies (i.e., center of activation, peak location, peak amplitude, spatial extent as function of amplitude). We will then compare these quantitative parameters across ROIs as a function of lobe (frontal vs temporal), hemisphere (left vs right), and scan (repeatability). We will address:

- 1) Claims: what are the test-retest repeatability coefficients achievable for center, spatial extent, and laterality of activation in language fMRI?
- 2) Conditions: what are the quantitative data quality constraints (e.g. SNR, performance, NVC, head motion) for meeting those claims?
- 3) Algorithms: which analysis and normalization strategies (e.g. AMPLE or alternatives) yield best repeatability?

This study will be carried out at two sites in parallel: Duke and Johns Hopkins, with each site using their own test-retest language fMRI data and carrying out their own analyses. Results from both sites will be compared and combined into the Language fMRI Profile (1.1) and for publication.

Deliverables

- 1) quantitative center, extent, and laterality claims for fMRI language Profile 1.1
- 2) quantitative data quality constraints to qualify fMRI Profile 1.1
- 3) representative language repeatability scans to support our findings will be uploaded to the QIDW

NUCLEAR MEDICINE COORDINATING COMMITTEE

FDG-PET Biomarker Committee:

A Procedure to Facilitate Greater Standardization of PET Spatial Resolution, PI: Martin Lodge, PhD – Johns Hopkins University

Project Description:

This project relates to a potential gap in the QIBA FDG PET/CT profile, namely that guidance regarding image reconstruction and spatial resolution is not yet adequately developed. PET spatial resolution directly affects standardized uptake values (SUV) so different sites could be QIBA-compliant, yet still produce tumor SUVs that differ substantially due to their use of inconsistent reconstruction parameters. We propose to develop an experimental procedure for measuring spatial resolution that could potentially be incorporated into future QIBA profiles as a tool to aid standardization of protocols across multiple sites.

The anticipated role of the proposed method is slightly different from the ongoing “harmonization” efforts being performed by US (Sunderland) and European (Boellaard) groups. These harmonization projects aim to

QIBA Proposed Workplan

determine reconstruction parameters and contrast recovery ranges that give rise to comparable quantitative results across scanner systems. Once harmonized reconstruction parameters have been defined, the organizers of multi-center studies will need to verify that potential sites have correctly implemented these recommendations. Specialist phantoms such as the NEMA image quality phantom are useful for determining harmonized parameters but are not well suited for widespread deployment. Phantoms with fillable compartments are notoriously prone to experimental error and many high quality centers do not have access to these phantoms nor the required experimental expertise. Site qualification that requires special phantoms of this sort is often associated with additional costs, delays and frustration, ultimately creating an unnecessary barrier to study participation. The current proposal describes a more practical experimental procedure that measures spatial resolution using a simple cylinder phantom of the sort that is already available at the overwhelming majority of sites. Cylinder phantoms are widely used for multi-center study qualification (e.g. ACRIN) and could potentially be employed in an expanded role, not just to assess scanner calibration accuracy, but also spatial resolution. It is anticipated that the proposed method will provide an essential verification tool that could be used to help confirm performance at individual sites and is therefore complementary to the ongoing “harmonization” efforts.

Conventional approaches for measuring spatial resolution involve point sources in air but this arrangement gives rise to unrepresentative results when images are reconstructed with the iterative algorithms that are typically used in clinical protocols. A requirement of the current application is that it should reflect the spatial resolution that is achieved with clinical protocols, not the limits of performance that can be obtained under optimized conditions. In addition, the method needs to be implemented uniformly and easily across all sites, including those with little or no physics support. We have previously proposed a method (Lodge et al. J Nucl Med 2009;50:1307-1314) based on measuring the extent of the blurring at the edges of a uniform cylinder phantom. This method has the advantages that it involves an extremely simple experimental procedure and a phantom that is already routinely used across the community for assessing calibration accuracy and image uniformity. Although the procedure worked well at our own institution, it did not transfer effectively to other sites. The problem was related to the need for higher spatial sampling than is usually used for whole-body applications. For the purpose of multi-center study qualification, sites were asked to reconstruct phantom images with 1 mm pixels. While not a major obstacle in theory, this requirement proved to be problematic for certain sites and made the method unsuitable for widespread deployment.

In light of this experience we developed a modified phantom procedure that did not require sites to change their clinical protocol in any way. Instead of positioning the cylinder phantom in the conventional orientation, parallel to the axis of the scanner, the phantom was positioned at a small angle, offset with respect to the z-axis (see figure). In this way the edge of the cylinder intersects the image matrix at slightly different positions in each slice. By appropriately shifting and then combining line profiles from different slices, the edge response function can be measured with very fine sampling despite the fact the original images may have a much wider pixel spacing, e.g. 4 mm. From these finely sampled, composite edge profiles we can determine the spatial resolution of the system by taking the first derivative, fitting a Gaussian function and measuring the full-width-at-half-maximum (FWHM). FWHM is a well-established metric for characterizing spatial resolution that is meaningful and intuitive for imaging specialists and non-specialists alike. While changes in FWHM at different points within the image are not captured with the proposed method, one of the biggest sources of resolution differences between scanner systems is the use of different post-reconstruction smoothing filters, a linear operation (typically) that can be well characterized using cylinder edge profiles.

Note the proposed method is applicable to a wide range of PET applications (including non-¹⁸F radiopharmaceuticals) and can play a role across multiple QIBA biomarker groups, e.g. FDG PET, amyloid PET & potentially SPECT.

Primary Goals and Objectives:

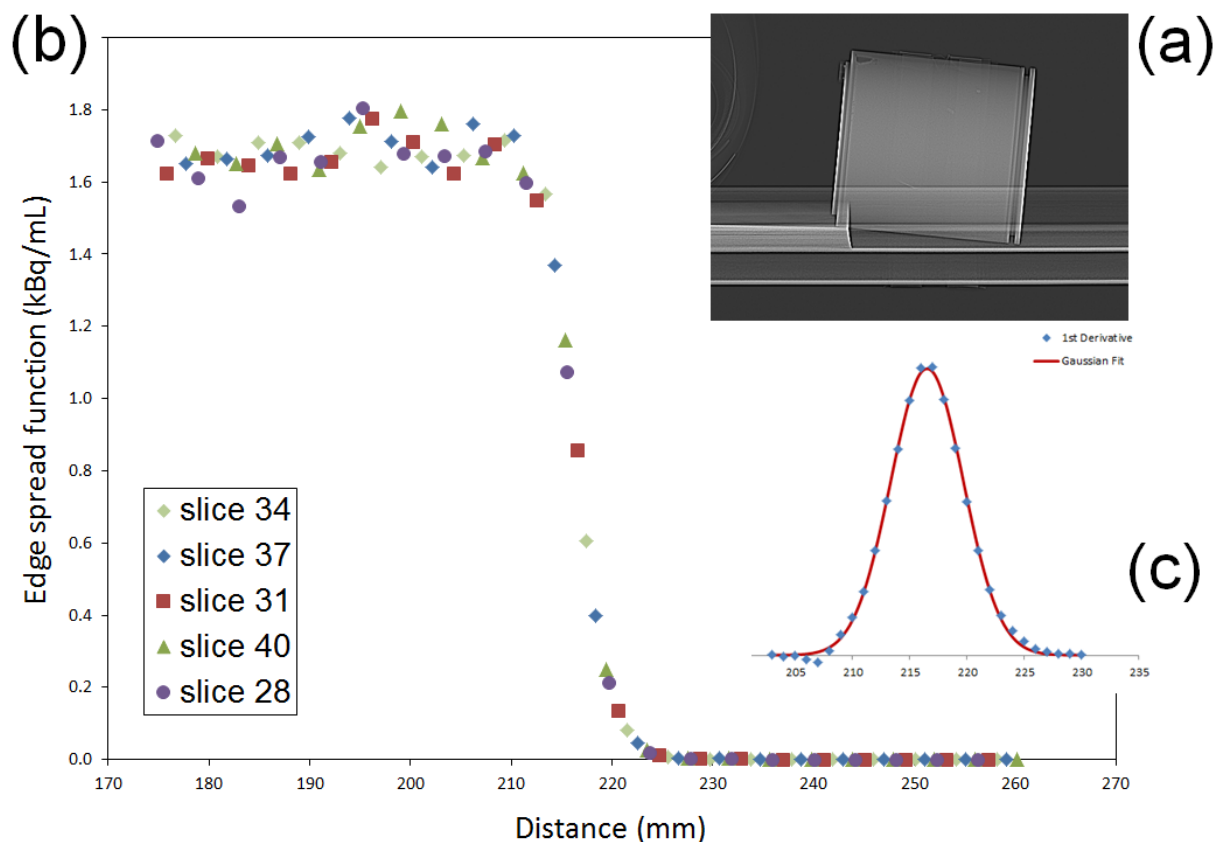
(1) Develop acquisition and analysis procedures for measuring the spatial resolution in PET images with uniform cylinder phantoms that are already available in the community. **(2)** Develop analysis software that is fully automatic and compatible with DICOM data from all major PET/CT vendors. **(3)** Test this procedure in a multi-site setting for whole-body and brain applications. **(4)** Study the relationship between spatial resolution measured using the proposed method and contrast recovery obtained using the NEMA image quality phantom for different sphere-to-background ratios. **(5)** Survey the spatial resolution that is being achieved with clinical

QIBA Proposed Workplan

protocols at participating QIBA sites. **(6)** Subsidiary aims include use of ^{68}Ge cylinders as a surrogate for ^{18}F ; compatibility with the ACR PET phantom; and potential use with SPECT.

Deliverables:

Reports at 3, 6 and 12 months. The final report will include a survey of the resolution achieved at QIBA sites and is expected to be augmented by freely distributable image analysis software.



(a) Lateral view showing a 20 cm uniform cylinder phantom positioned almost parallel to the axis of the scanner but with one end raised ~2 cm. **(b)** Multiple line profiles from different slices, each with 4 mm sampling, have been combined to form a more finely sampled edge spread function. **(c)** A Gaussian fit to the first derivative of the composite edge spread function can be used to measure FWHM.

Biologic and Reader Repeatability of FDG and CT Volumetric Parameters (ACRIN 6678 & MERCK), PI: Rathan Subramaniam, MD, PhD, MPH – Johns Hopkins University

Project Description:

The FDG PET and CT tumor volumetric parameters are valuable prognostic and predictive imaging biomarkers. There is no multicenter data on the biologic and reader repeatability of these parameters. This project investigates these properties of using the ACRIN 6678 and MERCK data that are held in ACRIN headquarters. Three readers (inter reader assessment) will perform segmentations and each reader will repeat the segmentations (intra reader assessment), > 90 days interval, in a random order, using threshold and gradient segmentations.

Primary goals and objectives:

1. To 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.

QIBA Proposed Workplan

2. To establish the intra reader and inter reader reproducibility and variability of MTV and TLG using arm C data of ACRIN 6678 and MERCK data
3. To establish the reproducibility using digital reference object (DRO) for volumetric parameters (MTV and TLG) – Collaboration with Paul Kinahan

Deliverables:

Manuscript on FDG PET volumetric parameter (MTV and TLG): Repeatability & Reproducibility

A PET Metabolic Tumor Volume Digital Reference Object (PET-MTV-DRO), PI: Paul Kinahan, PhD – University of Washington

Project Description:

Background

This project is a companion project to "Biologic and reader repeatability of FDG volumetric parameters (ACRIN 6678 & MERCK)" by Rathana Subramaniam, which is briefly summarized here as the metabolic tumor volume (MTV) repeatability study: The MTV-repeatability will use prospective NSCLC PET (stages III- IV) multi-center test-retest images from the ACRIN 6678 and Merck MK-0646-008 trials (74 patients total) to evaluate the reader repeatability of FDG PET tumor volumetric parameters, for which there is no multicenter data on repeatability. Three readers (inter reader assessment) will perform segmentations and each reader will repeat the segmentations (intra reader assessment), with at least a 90 day interval, in a random order, using threshold and gradient segmentations on clinical software.

Our study, the PET-MTV Digital Reference Object (DRO) study, will provide necessary extensions (i.e. features and procedures) to the FDG-PET/CT Digital Reference Object to expand the testing capabilities in order to provide ground-truth testing of image analysis packages that measure metabolic tumor volumes (MTVs) in PET. This project can support other QIBA projects evaluating FDG-PET MTV as a biomarker, such as test-retest or the construction of a Profile for FDG-PET MTV as a quantitative imaging biomarker.

Figure 1 shows a first pass at the PET-MTV-DRO, based on the NEMA Image Quality phantom, which has simple shapes and a homogeneous activity distribution.

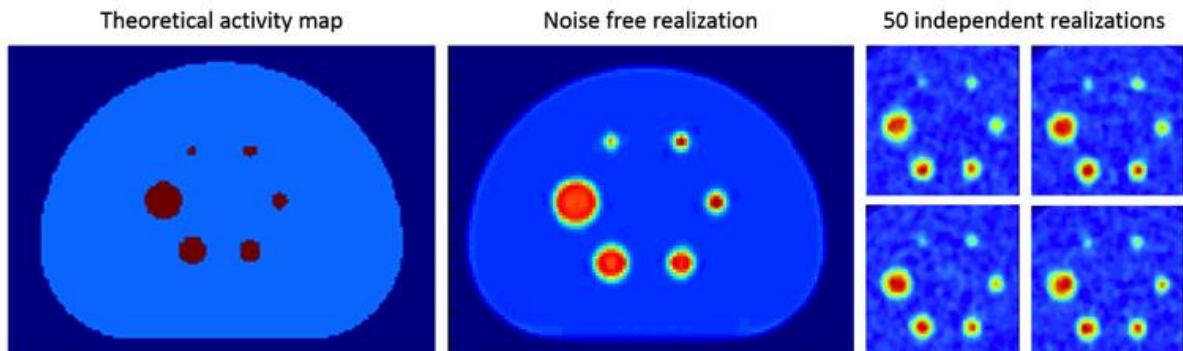


Figure 1. Left: A theoretically defined activity Digital Reference Object (DRO). Center: A reconstructed noise-free image (center) using our proven methodology to model PET/CT scanner physics to produce accurate resolution characteristics. Right: central regions (to show detail) of four of 50 noisy simulations and reconstructions of the DRO.

Figure 2 illustrates the concept of PET-MTV testing using the current FDG-PET/CT DRO. The FDG-PET/CT DRO is meant to test SUV calculations and is not appropriate for evaluating MTV segmentations as it has no resolution effects and is noise free. Nonetheless, it already illustrates important differences in the comparison of two widely used image analysis platforms.

QIBA Proposed Workplan

Project

The spherical objects in the first-pass PET-MTV-DRO shown in Figure 1 will be replaced with combinations of simple or complex shapes, each having simple or complex (heterogeneous) FDG distributions. The resulting images will have realistic levels of resolution loss (sometimes called partial volume errors) and noise levels appropriate for clinical whole-body FDG-PET/CT imaging. At the same time ground truth will still be known. Realistic PET images, noisy and noise-free, will be generated and stored in the proper DICOM format using methods we have developed as part of the original DRO project.

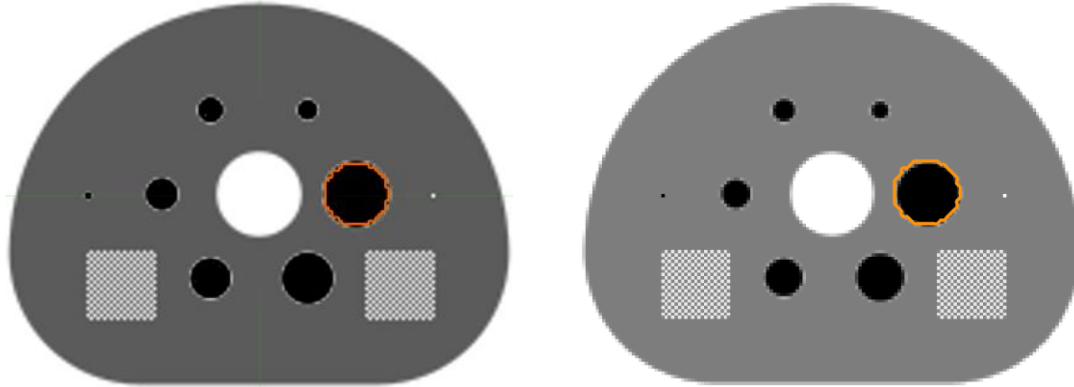


Figure 2: Screenshots of threshold-defined ROIs in MiM (left) and Osirix (right) for the largest sphere in the FDG-PET/CT DRO. The thresholds were defined using the built-in algorithms. MiM reports the volume of the largest sphere as 26.49 mL, while Osirix reports a volume of 29.13 mL, a 10% difference in reported volume.

Connection with MTV-repeatability study

The modified PET-MTV-DRO will be used as a calibration test as part of the MTV-repeatability study. The PET-MTV-DRO generation will be completed within the first month, before the first set of reader studies using patient images at the ACR Core lab. This allows a first set of testing by ACR Core lab and for readers to test a base level of operation of the ACR Core lab PET volume segmentation software. The MTV-repeatability study is targeting completion of the reader studies by the end of December of 2015.

Multicenter study

After the MTV-repeatability study, we will field-test other PET volume segmentation software packages at multiple sites using different display stations as we have successfully done previously. The primary site (University of Washington) will coordinate DRO distribution, testing procedures, and data analysis. The project will start with experienced quantitative imaging centers first, selected from QIBA FDG-PET/CT Technical Committee members, then expand to community centers if budget and time allow.

This will include a formal process description, publication of broad-spectrum testing results in a peer-reviewed scientific journal, potentially as part of the MTV-repeatability study, distributing announcements presentations and/or other advertising, providing a user manual and expected results and tolerance ranges, and providing a DRO check-out and result check-in process using the QIBA Quantitative Imaging Data Warehouse (QIDW).

Primary goals and objectives:

1. Updated version of the FDG-PET/CT DRO and user guide that includes additional features and procedures for testing metabolic tumor volume (MTV) measurement by PET/CT image analysis packages.
2. Incorporation of the PET-MTV-DRO into the MTV-repeatability study as both a training step and a reference / calibration check for ground truth and for comparison with other studies.

QIBA Proposed Workplan

3. Testing of the modified DRO at multiple member sites with a formal report of results and recommendations for use.
4. Potential inclusion with other QIBA projects leading to the development of FDG-PET MTV as a biomarker, such as multi-site test-retest studies using different MTV computation methods.

Deliverables:

1. Updated version of the FDG –PET/CT DRO and user guide that includes additional features for testing MTV software algorithms as well as recommended tolerance thresholds.
2. Tests of the modified DRO at multiple sites with a formal report of the results and recommendation for user.
3. Presentation and publication, as was for the FDG –PET/CT DRO.

PET Amyloid Biomarker Committee:

Analyses to Support Amyloid Imaging Profile Development, PI: Dawn Matthews – ADM Diagnostics, LLC

Project Description

The purpose of the proposed project is to address information gaps relevant to the development of image processing and analysis recommendations for the amyloid imaging profile. The first of these is the impact of subject motion. The second two are reference region definition and target region of interest definition, as evaluated under non-motion conditions and as related to subject motion effects. To provide this information, we propose to perform a set of motion simulations and reconstructions using de-identified florbetapir scans, analyze the resulting scans using variations in reference and target region definitions, evaluate the impact on test-retest variability, and put into context of other analyses we have performed and with the literature. We have identified, through a collaborative approach with Siemens and Avid Radiopharmaceuticals, the raw amyloid image data and reconstruction hardware and software platform necessary to accomplish these objectives.

Objectives and rationale

The specific objectives of this work are to:

1. Quantify the effect of subject motion, including misalignment between emission and transmission scan, on measured SUVR

Rationale: Subject motion is a major source of variability impacting the SUVR measurement that is the focus of the Profile claim(s). There are a limited number of studies on the effects of motion and motion correction; Mourik et al, 2009 and Ikari et al, 2012 are among the most informative and relevant. However, there is a gap in specific information for the amyloid target, likely processing and analysis methods, and context of use of the Amyloid Imaging Profile. A systematically generated set of data using a relevant amyloid tracer, applicable processing and analysis approaches, and controlled motion conditions would enable specific quality control recommendations and tolerance definition.

2. Quantify the effect of reference region selection, boundary definition, and spatial positioning upon measured SUVR

Rationale: A major influence (or confound) in SUVR repeatability is the reference region. Several published studies have made use of ADNI and other data to compare the group means and variance across different reference regions (Chen et al, 2015; Landau et al, 2015; Brendel et al, 2015; Matthews et al, 2014). Since different tracer manufacturers and protocols may recommend different reference regions, the Profile is not intended to prescribe use of a particular reference. However, in order to enable recommendations achieving the tolerances cited by the Profile Claim(s), it is important to extend this work to quantify impact at an individual level of variations in tissue selection, boundary definition, mask placement, and susceptibility to the motion effects studied in objective 2.1 above.

3. Quantify the effect of target region definition upon measured SUVR

QIBA Proposed Workplan

Rationale: A lesser, but still significant influence (or confound) in SUVR repeatability is target region definition. While some studies (e.g. Landau et al, 2014) have shown correlation between different overall methods (composites of target, reference, and boundary definition methods), there is a gap in quantifying the impact of well-defined variations in target region definition at the individual level, including susceptibility to motion. Filling this gap would enable specific Profile recommendations.

Deliverables:

1. Quantify the effect of misalignment and subject motion: This prospective evaluation will make use of raw amyloid PET data selected from a set of 45 florbetapir scans (15 healthy controls (HC), 15 MCI, 15 AD) owned by Avid and residing at Siemens. This data was acquired for each subject over a period of 10 min in list mode on a Siemens mCT system from the University of Tennessee in Knoxville, enabling great flexibility in rebinning for time frames. We will select a total of twelve scans comprised of three to four scans from each of the following three categories: amyloid negative HC, amyloid positive MCI with cortical average SUVR near threshold (as available), and AD with high amyloid burden. For each category, we will target those scans having the least intrinsic motion, and representing a range of brain morphologies, skull thicknesses, and tissue-skull gaps. In addition, if subjects having a “halo” (hyperintensities exterior to cortex) are identified as is found in some florbetapir scans, these subjects will be included for comparison.

We will evaluate: (1) the effect of misalignment between emission and transmission (CT) scan, (2) the effect of within-emission scan subject motion where emission-transmission scan alignment is not corrected on a frame by frame basis, and (3) the effect of within-frame motion that may not result in an overall detected between frame mismatch (e.g. subject returns to position). In addition, we will evaluate the effects on cortical average SUVR across different reference regions and different approaches to target region definition (aspects of 3.2 and 3.3). Since there is a vast combination of translational and rotational combinations, we will define a series of translations and rotations that in combination address the spectrum of likely motion cases. Additionally, we will examine an alternate (lower) simulated scanner resolution achieved through re-slicing and smoothing in a subset of data. For implementation, a protocol and analysis plan will be developed to define scan selection criteria, the translational and rotational permutations, time frame duration, whether or not PET-CT alignment is corrected for each frame, processing methods, target and reference region definitions, and statistical analysis. To enable measurement of the impact of emission-transmission alignment, for each translation and/or rotation, Siemens will transform the mu-map generated from the CT according to the pre-determined parameters and then reconstruct the image. For evaluating the impact of subject motion without frame-specific correction of PET-CT alignment, prescribed translations and rotations will be performed on the reconstructed images, followed by SUVR measurement. ADMdx will process and analyze all image data.

The deliverable will be a report detailing the quantitative impact on measured SUVR for the various conditions, across the three categories and with attention to impact of variables such as skull thickness or tracer halo presence. This may lead to a published manuscript on behalf of QIBA.

2. Optimal reference region: We will integrate published findings with additional results of processing and analysis of ADNI florbetapir images using several different reference regions. We will examine the impact of boundary definition (e.g. inclusion/exclusion of all slices for cerebellum, erosion away from neighboring tissue), and placement errors. This will provide a level of individual subject error impact and insight to reference region impact beyond published data. Deliverable will be a report identifying the range of contributions from each potential variable examined to overall measurement variance as related to the tolerances specified in the Claim(s).

Effect of target region definition: We will compare SUVR results using the various typically applied target region definition approaches (regions selected to contribute to cortical average, all or portion of region, gray masked vs. not, eroded vs. not), the impact of region placement, and susceptibility to motion. Deliverable will be a report identifying the contributions from each potential variable examined to SUVR variance as related to the tolerances specified in the Claim(s).

QIBA Proposed Workplan

Amyloid Brain PET Test-Retest Meta Analysis, PI: Rathana Subramaniam, MD, PhD, MPH – Johns Hopkins University

Project Description:

This is a systematic review and meta-analytical project on test – retest repeatability of ^{18}F Amyloid brain PET radiopharmaceuticals to establish the claim for the Amyloid brain PET profile.

Primary goals and objectives:

To establish the repeatability co-efficient and coefficient of variation for ^{18}F Amyloid brain PET radiopharmaceuticals to establish the claim for the Amyloid brain PET profile

Deliverables:

1. Systematic review of PubMed, EMBASE, Cochrane data bases and perform meta analysis to establish the repeatability co-efficient and coefficient of variation for longitudinal claim
2. Systematic review and meta-analytical manuscript

ULTRASOUND COORDINATING COMMITTEE

SWS Ultrasound Biomarker Committee:

Analysis of Sources of US SWS Measurement Inter-System Variability, PI: Mark Palmeri, MD, PhD – Duke University

Project Description:

There have been two primary foci for the System Dependencies & Phantom Development US SWS subcommittee during the currently active Round IV funding:

1. Phase II viscoelastic (VE) phantom development and correlation with *in vivo* liver data, followed by measurements at different academic, government and industry sites on a variety of US SWS platforms, and
2. Development, validation and distribution of digital phantoms that simulate shear wave propagation in elastic and viscoelastic media for a variety of transducer and material properties.

The VE properties of the Phase II phantom recipes from CIRS, Inc. were characterized using two metrics: (1) phase velocity at 200 Hz, and (2) a linear dc/df slope ranging from 100-400 Hz. These empirical metrics avoided having to rely on a (overly simplified) material model that might not be appropriate for liver tissue. External material testing was done at the Mayo Clinic, and comparison with *in vivo* liver data was done at Duke (Figure 1(a)) and Philips (using data acquired at the Mayo Clinic). The Phase II phantoms have been measured at over half of the sites to date (Figure 1(b)).

QIBA Proposed Workplan

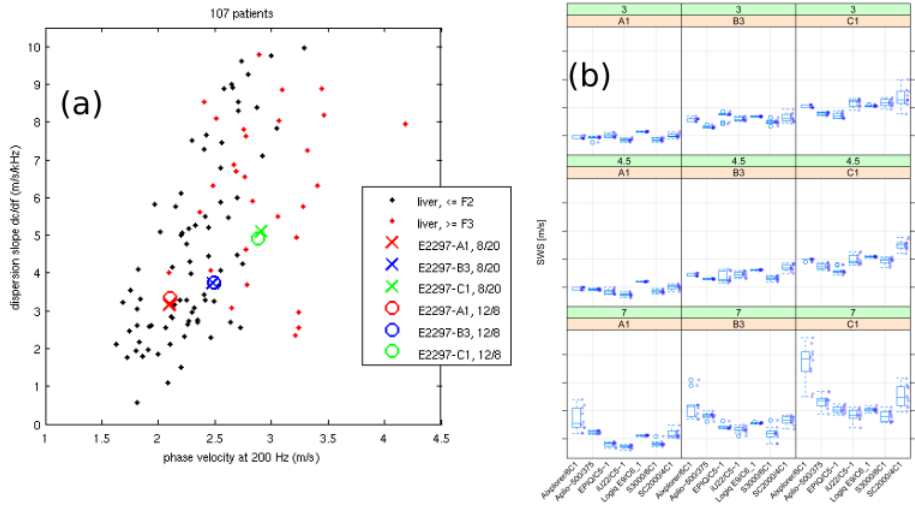


Figure 1: (a) CIRS phantom measurements (circles and crosses) made with a research SWS imaging system on different dates compared to human data categorized by fibrotic disease severity. (b) Intersystem variability across different phantoms [A1, B3, C1] and imaging depths (3, 4.5, 7 cm). Absolute SWS values have been withheld due to the ongoing, blinded nature of the study.

In parallel to the Phase II phantom studies, Duke and Mayo Clinic have been developing and validating Finite Element Method shear wave simulations to generate digital phantoms of shear wave propagation from different curvilinear transducer configurations in elastic and viscoelastic media. The source code for this simulations is publically available on GitHub (<https://github.com/RSNA-QIBA-US-SWS/QIBA-DigitalPhantoms>), and the datasets are available on QIDW (<http://www.rsna.org/qidw/>). These simulations have been validated across two commercial FE solvers—LS-DYNA and Abaqus—to demonstrate the absence of numerical bias (Figure 2), and the datasets also have corresponding processing code and example outputs provided in the GitHub repository to allow for in-house validation with other FE solvers. These digital phantoms will allow manufacturers to evaluate sources of bias in their shear wave reconstruction algorithms as a function of focal configuration and material stiffness and viscosity, all of which have been identified as potential first-order sources of inter-system variability.

Group Velocity (Time of Flight, Z = 70 mm)

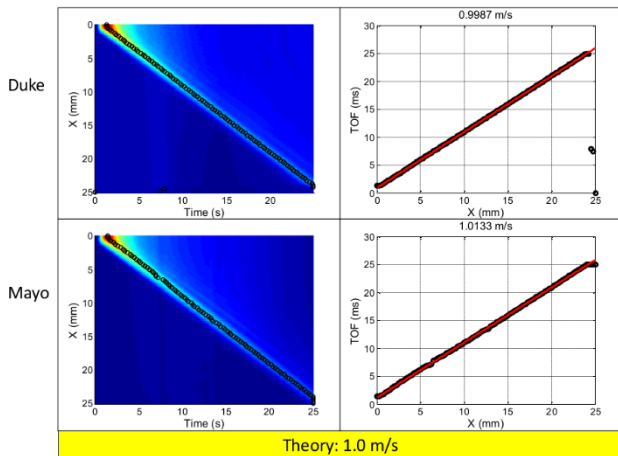


Figure 2: An example comparison of the propagating shear wave displacement field at the focal depth (70 mm) for a F/3.5 curvilinear focal configuration in a $E = 3$ kPa elastic medium with a $334 \mu s$ excitation duration. The estimated group shear wave speeds for the LS-DYNA (Duke) and Abaqus (Mayo) solvers are within 0.015 m/s of the each other and the theoretical wave speed of 1 m/s.

All of our Round 4 funding objectives will be completed once all of the viscoelastic material configurations corresponding to the Phase II phantoms have been simulated and uploaded, which we do not foresee any delays in completing.

Primary objectives, deliverables and timeline:

As shown in Figure 1(b), the Phase II phantom data are showing intra- and inter-system variance as a function of focal configuration and VE material. We know that different focal configurations yield shear waves with different spectral content, and viscosity yields shear waves that propagate at different phase velocities. We also know that different focal configurations and measurements at depth away from the focal depth can

QIBA Proposed Workplan

introduce sources of shear wave speed reconstruction bias.* Given our Phase II data and these potential sources of bias, we proposed to following objectives for Round 5 funding:

Objective (Institution)	Motivation	Deliverable (Due)
1. Perform statistical analysis on complete Phase II phantom data from all measurement sites. (Duke)	Quantify the inter-system bias and intra-system variance in our most calibrated materials to date. These findings will be correlated with the shear wave spectral content associated with the different focal configurations and VE materials, which will be expanded in simulation in Objective 3.	<ul style="list-style-type: none"> a. Compile data in spreadsheet posted to wiki. (Q2) b. Compile report of statistical analysis to distribute to group. (Q3) c. Prepare manuscript. (Q4) d. Present data at RSNA, IEEE, etc. (Q4) e. Integrate findings into the profile. (Q4)
2. Expand the digital phantoms to include a greater variety of focal configurations that would generate shear waves with different spectral content. (Mayo)	The Phase II and digital phantoms to date have been restricted to a small set of curvilinear focal configurations. This objective will expand these focal configurations to establish upper bounds of variability.	<ul style="list-style-type: none"> a. Establish focal configurations to simulate. (Q1) b. Run simulations; compile and analyze data. (Q2) c. Upload datasets to QIDW. (Q4) d. Prepare report on findings, along with related manuscripts, presentations. (Q4)

* Zhao, et al. "Bias observed in time-of-flight shear wave speed measurements using radiation force of a focused ultrasound beam," UMB, 37(11):1884-1892, 2011.

QIBA Proposed Workplan

<p>3. Augment the Round IV elastic and VE digital phantoms to include ultrasonic displacement estimation that captures bias and jitter artifacts that can confound SWS estimation. (Duke)</p>	<p>One primary difference between the digital and Phase II phantom data is the displacement underestimation and jitter associated with ultrasonically tracking the shear wave displacement fields.[†] This objective will introduce these artifact sources into the digital phantoms for another dimension of SWS variance analysis using existing tools.[‡]</p>	<p>a. Establish focal configurations to simulate. (Q1) b. Run simulations. (Q2) c. Compile and analyze data; upload datasets to QIDW. (Q3) d. Prepare report on findings, along with related manuscripts, presentations. (Q4)</p>
---	--	---

[†] Palmeri et al. "Ultrasonic tracking of acoustic radiation force-induced displacements in homogeneous media," IEEE UFFC, 53(7):1300-1313, 2006.

[‡] <https://github.com/mlp6/ultratrack>

QIBA Proposed Workplan

<p>4. Evaluate the impact of modulated spatial acoustic radiation force source distribution from attenuation and phase aberration on SWS reconstruction. (Mayo)</p>	<p>A confounder that most likely exists <i>in vivo</i> but is not captured in our current digital phantoms or the Phase II phantoms is distortion of the acoustic radiation force distribution that could violate the assumptions of the time-of-flight algorithms used to estimate SWS.</p>	<p>a. Develop simulations with phase aberrating superficial layers. (Q1-2) b. Run simulations. (Q3) c. Compile and analyze data. (Q3) d. Prepare report on findings, along with related manuscripts, presentations. (Q4)</p>
---	--	---

CROSS-MODALITY PROJECT

Aggregated Measures of Agreement for QIB Validation: An Open Source Toolkit, PI: Daniel Barboriak, MD – Duke University

Project Description: The purpose of this project is develop open source software to calculate aggregated measures of agreement in order to facilitate image analysis algorithm development, comparative analysis of algorithm output, and demonstration of technical compliance. This project leverages activities in the QIBA metrology effort, development of the QIBA DCE-MRI DROs and work on the open source QIBA DRO Evaluation Tool (QDET, Round 3 Project 9) to provide a toolkit which will support image analysis algorithm verification, a critical step in quantitative imaging biomarker (QIB) validation.

Primary goals and objectives

When different image analysis algorithms are tested against purported ground truth, they are generally found to derive QIBs with different levels of bias and precision. As was noted in the QIBA metrology effort, in this situation it may become challenging to determine which method has superior performance to another, and which methods have achieved adequate levels of performance to be useful in achieving the goals of the QIB profile. Not infrequently, if an algorithm is tailored to maximize precision, biases may be introduced, and vice-versa. It would be helpful to have simple combined aggregate figures of merit to simplify algorithm evaluation. **One barrier to this process of evaluation is that there is not easy access to these aggregated measures of statistical agreement, nor widespread facility in interpreting the results of these measures.**

In order to address this specific barrier, we propose to develop an open source toolkit with the following characteristics:

1. Using the open source QDET package as a starting point, provide the following statistical analyses, verified against existing software packages:

QIBA Proposed Workplan

- 1.1 Concordance Correlation Coefficient (CCC)
 - 1.2 Root mean square deviation (RMS)
 - 1.3 Total Deviation Index (TDI)
 - 1.4 Bland-Altman limits of agreement (LOA)
 - 1.5 Sigma analysis based on estimates of allowable total error (<http://www.ncbi.nlm.nih.gov/pubmed/24615486>). This measurement may be of interest because it places QIB algorithm evaluation into a similar context as used for clinical laboratory performance assessment, a context which may be more understandable by FDA.
2. Allow inputs of imaging result / DRO ground truth (implemented in QDET) or input of matrices of means, squared errors and nominal ground truth to be used in the evaluation.
 3. Develop, if statistically valid, unequal weighting of different ground truth parameter combinations within the statistical analyses.
 4. Provide, for clearly defined evaluation scenarios, written guidance to the interpretation of the various statistical analyses.

It is important to note that this toolkit is designed to be valuable in the evaluation of not only the QIBA DCE-MRI DRO, but also many of the growing number of DROs being developed in the QIBA portfolio developed using an assumed ground truth; for example, QIBA FDG-PET DROs. **By allowing matrices of nominal ground truth, means and squared errors to be used as input, this toolkit is also designed to extend t analysis using aggregated metrics to results derived from physical phantoms such as QIBA DWI phantom, T1 response phantom, and lesion simulation phantoms.**

Deliverables

1. Provide open source access to statistical analyses.

A revised open source package, based on the Python source code for the QDET project, which will provide the following aggregated methods for measurement of agreement: CCC, RMS, TDI, LOA and Sigma. For CCC and RMS, we will verify pre-existing QDET code, and for TDI and LOA we will produce code and verify against external packages such as R (<http://www.r-project.org/>). The figures of merit will be provided as part of GUI or .csv files for use in further analysis.

2. Demonstration of use of aggregated measures of agreement to tune image analysis algorithm parameters.

The toolkit will be used in a demonstration project to tune software algorithm parameters to maximize figures of merit. We propose using the jsfim function in the modelling software package JSim as applied to curve fitting to perform T1 mapping (http://www.physiome.org/jsim/docs/User_Fim.html). One advantage of using JSim is that it allows a number of curve fitting solvers, each with several parameters such as step size and maximum iterations, to be compared. An online document will be created to describe the experimental set up and to provide the scripts to work through the example.

3. Demonstration of use of aggregated measures of agreement to rank performance of competing image analysis algorithms.

The toolkit will be used in a demonstration project to compare software performances using a variety of metrics, including the aggregated metrics above. We propose using the pre-existing outputs for T1 mapping performed by approximately 10 software analysis algorithms that we evaluated in the DCE-MRI DRO project, and create a scoreboard of ranks by parameter. Again, an online document will be created to describe the experimental set up and to provide the scripts to work through the example.

4. Evaluate the performance of aggregated measures of agreement, and provide guides to interpretation.

This deliverable will be performed with the assistance of Nancy Obuchowski, Ph.D. of Cleveland Clinic Medical Center. The performance of various aggregated measures of agreement will be critiqued, and written guidance to the interpretation of these metrics will be provided as part of the software toolkit output.

QIBA Proposed Workplan

5. Allow unequal weighting of parameter combinations in evaluation of image analysis algorithms.

Heterogeneous weighting would allow combinations of ground truth parameters considered to be more important (for example, by consensus of experts, by understanding of the pathophysiological impact of particular QIB values, or by pre-existing knowledge of parameter combinations most frequently encountered in clinical use) to have an appropriately greater impact on the results of the analysis. Statistical input from Nancy Obuchowski will also support this deliverable.