Quantitative Imaging Biomarkers Alliance (QIBA)

FINAL REPORT: AS OF MARCH 31, 2013

EXECUTIVE SUMMARY

With support of the now completed contract, QIBA is now known and visible as a collaborative, multidisciplinary infrastructure to foster research, approval, and use of quantitative imaging biomarkers.

During the contract term, QIBA has converged its process for the development, validation, qualification and use of accurate, repeatable quantitative imaging biomarkers across instruments and settings. The paradigm that has evolved is to create, for each imaging biomarker, documents referred to as the UPICT Protocol and the QIBA Profile. Definitions of these documents can be found on the QIBA wiki at http://qibawiki.rsna.org/index.php?title=What_Are_Profiles_and_Protocols%3F#Definitions_and_Descriptions_for_QIBA_Profiles_and_Protocols. To create these documents, QIBA coordinates broadly with various stakeholders, including professional imaging societies, academic centers, imaging device manufacturers, the pharmaceutical industry, and federal agencies. Profiles have been released for CT Volumetry and DCE-MRI, with FDG-PET SUV in the final stage of addressing public comment. Profiles for Lung CT densitometry and airway measurements for COPD, CT imaging in the setting of lung cancer screening, and fMRI for pre-surgical mapping are currently being developed.

We have also conducted 26 projects funded through this contract to further improve the accuracy and precision of imaging biomarkers, pursuing needed research regarding sources of bias and variability and the impact of various proposed mitigation strategies. Final reports on all those projects are summarized below, as well as being included in full text in attachments.

We have partnered with the FNIH Biomarkers Consortium, completing detailed Briefing Documents for FDG-PET and CT Volumetry, meeting with the FDA's Biomarker Qualification Review Teams, and submitting formal revisions updating with results from the various QIBA studies.

This report is stated in terms given in the accepted Work Plan. Work Tasks 11-19 in the SOW are associated with the overall program. Some of these do not lend themselves to scheduling in Gantt chart format, but are associated with specific events during the contract term. A final report on those activities is given in Section A below.

Work tasks 1-10 generally comprise the elements of our roadmap for each biomarker. Section B lists specific experimental groundwork projects that are ongoing and/or have been allocated NIBIB contract project funding. Sections C-F provide a high-level statement of what has been accomplished for CT, FDG-PET, MRI, and our new initiative in Ultrasound respectively.

A. OVERALL PROGRAM TASKS 11-19.

Our initial plan with respect to each numbered task is given in italics, and the status is given in BOLD font:

NIBIB Task 11.   Stimulate an interest in disseminating and implementing QIBA solutions to assess their feasibility and efficacy more broadly.

a. *We will schedule two QIBA meetings per year, one in May and the other at the RSNA Annual Meeting in November, with agenda set for this purpose.*

Well-attended public and working meetings have been held during the RSNA Annual Meeting in December 2010, November 2011, and December 2012. Formally scheduled sessions included Special Interest Session - Quantitative Imaging Biomarkers for Clinical Care and Research and a Technical Committees Working Meeting. We have held annual QIBA meetings as well in May 2011 and 2012. We also arranged and conducted six vendor meetings (Siemens, GE, Philips, Toshiba, and post-processing software suppliers in both CT Volumetry as well as FDG-PET).

b. *We will schedule educational content in the RSNA Annual Meeting to disseminate information to a wide audience.*

At each RSNA Annual Meeting, our technical committees prepared and presented a poster associated with our QIBA Kiosk and associated "Quantitative Imaging Reading Room" exhibit. General updates were provided each year in a Special Interest Sessions.

c. *We also publish a QIBA Newsletter electronically.*

During the contract term, we regularly published the "QIBA Newsletter." Examples of featured articles include "Standardization of Clinical Trial Image Acquisition is Essential for Establishing Clinical Utility" by GARY S. DORFMAN, MD; "Challenges in Tackling Quantitative DCE-MRI" by ALEXANDER GUIMARAES, MD, PhD.; "How QIBA Will Benefit Medical Device Innovation," by SANDEEP N. GUPTA, PhD; "Software Development for Analysis of QIBA DCE-MRI Phantom Data," by EDWARD ASHTON, PhD; reviews of RSNA 2011: Quantitative Imaging/Imaging Biomarkers and QIBA Meetings and Activities; and a PubMed Search on How QIBA Will Benefit Medical Device Innovation.

NIBIB Task 12.   Encourage adoption, integration and clinical education of validated QIBA solutions by the research and industry community.

a. *We have begun to schedule company-specific meetings with managers of medical device companies to explain QIBA, and solicit their feedback.*

Our technical committees have had broad involvement from the supplier community, including medical device as well as software companies. Building on the meetings we held in the first contract year with manufacturers, we have held a discussion to bring together the 3rd party manufacturers of display stations for FDG-PET (with discussion among value-added software companies largely having been driven by CT analysis companies) to present the QIBA perspective and why the Profiles matter to them. We also made a presentation to the Nuclear/PET MITA meeting about QIBA, and are planning on doing so again this year.

In FDG-PET and DCE-MRI, synthetic data objects ("digital reference objects", or "DRO"s) have been created and are being evaluated by a number of companies. For example, the FDG-PET DRO has been used by the following:

- **GE**
- **Hermes**
- **Keosys**
- **MedImage**
- **MIM**
- **OsiriX**
- **Philips**
- **PMOD**
- **Segami**
- **Siemens**

In CT Volumetry, a pivotal algorithm performance "Challenge" has been conducted with the following participants (the pilot having concluded earlier with a partially overlapping group of participants):

- **Median Technologies**
- **Fraunhofer Mevis**
- **Siemens**
- **Toshiba**

- **Icon Medical Imaging**
- **Columbia University**
- **INTIO, Inc.**
- **NYU Langone Medical Imaging Center**
- **Perceptive Informatics**

In addition to the vendor meetings at RSNA Annual Meetings, we have completed on-site visits with CT, NM, and MR business units at GE in Waukesha, WI; with the MR business unit at Philips in Cleveland, OH (with video conference feed to Best, NL); the MI business unit at Siemens in Knoxville, TN; and the Toshiba Research Institute-USA with representatives from MR, NM, and CT.

b. *We will work with the Pharma Imaging Group (PIG) to get QIBA solutions integrated into pharmaceutical industry drug trials.*

During the contract term, we completed the field test of our first controlled document, a protocol for CT volumetry in lung cancer. We hold regular status discussions with the Pharma Imaging Group (PIG), and the PIG chair is a member of the QIBA Steering Committee.

We worked especially closely with the PIG in the evaluation and collection of comments in response to the FDA's public comment period for "Standards for Clinical Trial Imaging Endpoints." We solicited and collated comments from the QIBA membership as well as produced a matrix with detailed remarks on paragraphs within the guidance and observations as well as suggestions from the vantage point of our experience with UPICT protocol and QIBA profile authorship.

c. *We will work with ACRIN, the SNM Clinical Trials Network, and other academic organizations to get QIBA solutions integrated into clinical trials.*

We have refined the process for the Uniform Protocols for Imaging in Clinical Trials (UPICT) initiative to discuss details for specific consensus protocols. The most active project in this regard is working towards a consensus protocol for quantitative FDG-PET and we also have made progress with an Alzheimer's protocol.

NIBIB Task 13.   Develop an initial consensus on quantitative imaging biomarkers qualification by coordinating broadly with various stakeholders, including professional imaging societies, academic centers, imaging device manufacturers, and drug industry.

a. *We will use breakout groups at the annual "Imaging Biomarkers Roundtable" to achieve this objective, as well as collective input from the Pharmaceutical Imaging Group, meetings with individual medical device manufacturers, and recommendations from relevant academic workshops.*

We have matured our processes, particularly those associated with review and release of QIBA Profile documents. We have formalized a process to disseminate program documents, including an inclusive list of professional societies, trade groups, and other organizations. We have begun to use this process to engage stakeholders in the review of our documents through the public comment process. We have also made substantial progress on the meaning and actualization of compliance. This continues to be an active area of development for us.

Additionally, we have now formed an active Ultrasound Modality Committee, which is currently working on shear wave speed (SWS) as a quantitative biomarker. We have also

been responsive to the feedback we have heard on our current biomarkers and proposals for new ones. For example, we have formed the effort for DWI-MRI which has been organized associated with our DCE-MRI effort.

We held specific discussions with these groups associated with our September 2010 Imaging Biomarkers Roundtable, with targeted break-out sessions considering neurological diseases and regulatory issues related to contrast agents (in addition to the prior subjects we have raised in that meeting). Additionally, we invited and received reports from the FDA as well as the Critical Path Institute, and also featured a discussion regarding Ultrasound. Since that meeting we have reached out to several key opinion leaders to discuss creation of an Ultrasound effort as well as to potentially expand the MRI effort to address Diffusion Weighted Imaging.

b. *For consensus related to formal FDA qualification of imaging biomarkers, we will work with the FNIH Biomarkers Consortium and the Critical Path Institute as well. This collaboration will occur by monthly conference calls, as well as collective work on the Briefing Documents and Data Packages to be submitted to the FDA.*

During the first contract year, we prepared and submitted 100+ page briefing documents to FDA for quantitative FDG-PET as well as CT Volumetry. We have prepared for and met with each respective Biomarker Qualification Review Team (BQRT). Subsequently we completed substantial updates to the Briefing Documents for FDG-PET and CT Volumetry, submitting a formal revision on the latter to the FDA in September.

We have also continued our engagement of potential data donors, presently including active discussions with Merck, Astra Zeneca, and Genentech/Roche. An approach whereby we prepare and submit quarterly updates to the Briefing Documents highlighting additional study designs and data sets has been assembled.

NIBIB Task 14.   Organize and manage relationships in a collaborative, multi-disciplinary environment that fosters communication among imaging groups and other medical disciplines involved in the research, approval and use of quantitative imaging biomarkers.

a. *The QIBA Steering Committee meets once per month by phone and in person twice a year.*

We have held monthly teleconferences of the Steering Committee as well as face-to-face meetings at multiple times throughout the contract term. Discussions included the specific administration of NIBIB project funds as well as all other matters associated with completion of the contract work items.

b. *The Modality Committees convene on an as-needed basis.*

We focused the modality committees on evaluation of funding requests and the recommendations to the Steering Committee on funding.   We also used them for a deliberative voting process by committee members as to readiness for the Profile documents. We have formed a new Modality Committee for Ultrasound, generalized the MRI work to include "perfusion, diffusion, and flow" (PDF), and started a new CT Volumetry initiative in Lung Nodule Assessment in CT Screening. Each of these has been approved by the Steering Committee.

c. *The Technical Committees meet biweekly, with groundwork subgroups meeting as needed, often weekly. All of these QIBA groups are composed of individuals from the named stakeholder groups.*

Each of the teams, or subgroups, meet by phone at least biweekly, and many meet more frequently.

NIBIB Task 15.   Create and implement a process by which standardized and harmonized systems emerge that are sufficient for the development, validation, qualification and use of accurate, repeatable quantitative imaging biomarkers across instruments and settings.

The QIBA Steering Committee, with input from the Technical Committees, has begun to develop such processes. These will be documented in a process manual by the end of year 1 (Sept 30, 2011). We will provide a feedback (public comment) mechanism with a formal update mid-way through year 2 (March 30, 2012).

Profile Development and Release Stages: The Technical Committee Profiles are at different stages of completion, which needs to be recognized, but then harmonized in the QIBA Process. A similar process was employed by IHE. A proposed nomenclature for classifying the evolutionary stage of Profiles based on the intensity of input, testing and consensus achieved to date has been approved by the Steering Committee.

Metrology Terminology and Methods: We have addressed challenges deriving from lack of standardized measurement terminology and methods used by QIBA Technical Committees and the imaging field in general by convening a QIBA Metrology Workgroup, which has held two face to face workshops as well as a large number of teleconferences. Groups were formed to address Terminology, Technical Performance Characteristics and Algorithm Comparisons. Varying stages of progress have been made but all groups agreed that the resulting guidelines will be valuable to the QIBA Technical Committees and should be implemented into the QIBA Process. We have conducted many weekly call sessions across three workgroups of this initiative in converging consensus metrology terminology and statistical approach. Five manuscripts are in late stages of development on these topics as well as examples of them.

Harmonizing Profile Language: Profile editors have been working together to consider uniform format and terminology. Specifically with respect to Claims, the Metrology Workshop participants will provide input on approaches that would provide uniformity.

We reviewed and consolidated process and structure including authoring, review, and testing of documents; imaging biomarker roadmap and approach to groundwork; as well as team structure & governance. We also compiled and reviewed various models for QIBA compliance.

In addition to the Metrology Workgroup, we formed a second *ad hoc* workgroup to address key issues associated with our work. This, QIBA/ Radiology Informatics Committee (RIC), converged proposals to form an image warehouse with associated capability that is presently in a pilot stage and is being tested for functionality. This latter workgroup utilized input derived from the NIBIB funded project to engage the technical committees on requirements and issues associated with data resources.

During the first contract term, a preliminary version of the Process Manual has been posted to the QIBA Wiki and different steps of the process are being refined through experience by the various teams. Also in the contract term, we created means for sharing content between protocols and Profiles as well as refined formats and writing methods.

NIBIB Task 16.   Clarify and optimize the regulatory pathway by which quantitative imaging biomarkers enter the market.

During the first contract term, two Special Reports have been published in Radiology:

Article "A Collaborative Enterprise for Multi-Stakeholder Participation in the Advancement of Quantitative Imaging" has been published by Radiology. This paper is available online at http://radiology.rsna.org/cgi/content/abstract/258/3/906.

Article "Quantitative Imaging Test Approval and Biomarker Qualification: Interrelated but Distinct Activities" has been published by Radiology. This paper is available online at http://radiology.rsna.org/cgi/content/abstract/radiol.10100800.

a. *We have also initiated formal efforts with FDA/CDER to qualify two biomarkers utilizing these ideas as of this year. We expect to meet with the FDA in a collaborative process and then transition to the formal review phase. As these processes are new to both FDA and us, we are not able to indicate a schedule at this time but will update in our periodic reports. Additionally, early in Year 2, we anticipate formal discussions related to the use of data accumulated for qualification to be contributory to CDRH filing and will update as we get closer to that engagement.*

As mentioned above (Task 13a), we met with the BQRT for each of the two imaging biomarkers on which we seek a qualification result and have started to strategize how to complete the Full Data Packages for them.

As detailed above under Task 13, we have continued the engagement with FDA regarding biomarker qualification. Also during the second year we produced a detailed response to the FDA's public comment period for "Standards for Clinical Trial Imaging Endpoints." In the last 6 months, we have completed projects which produced data capable of being used directly in regulatory pathways. We are also now providing additional specific details on how to establish compliance with QIBA profiles that may use data sourced and/or curated by QIBA efforts.

As detailed above under Task 13, we have corresponded with the FDA regarding questions they raised, questions we have raised, as well as the process and logistics for updating the briefing documents and reconvening the BQRTs to review them.

Additionally during the current contract term, we produced a detailed response to the FDA's public comment period for "Standards for Clinical Trial Imaging Endpoints." We solicited and collated comments from the QIBA membership as well as produced a matrix with detailed remarks on paragraphs within the guidance and observations as well as suggestions from the vantage point of our experience with UPICT protocol and QIBA profile authorship.

NIBIB Task 17.   Establish a process for relating biomarkers to disease areas, setting the clinical context and, based on the clinical context, identifying and prioritizing what biomarkers to pursue.

We will use breakout groups at the annual "Imaging Biomarkers Roundtable" to achieve this objective.

In addition to what has already been noted regarding specific topics and breakouts at the Imaging Biomarker Roundtable (Task 13b), the Steering Committee has developed criteria and a procedure for considering and prioritizing new biomarkers to address.

Our new Modality Committee for Ultrasound is now active, the MR Modality Committee has actualized its extension for other biomarkers beyond DCE, and the CT Volumetry committee has extended its effort to the context of Lung Nodule Assessment in CT Screening.

During the current contract term, we have formed a new Modality Committee for Ultrasound, and the MR Modality Committee has proposed an extension of their work to include DWI-MRI that has been approved by the Steering Committee.

NIBIB Task 18.   Create a collaborative, multidisciplinary infrastructure to foster research, approval and use of quantitative imaging biomarkers, including development and maintenance of a national repository of quantitative imaging biomarker data, representation at a variety of workshops and meetings, and provide project management and staff support for same.

a. *The QIBA committee structure and leadership constitutes one component of a collaborative, multidisciplinary infrastructure to foster research, approval and use of quantitative imaging biomarkers. A plan for long-term sustainability will be developed over the next year. (See Task 19).*

A panel of experts, convened by the RSNA Board of Directors and chaired by Carolyn Meltzer, MD, Emory University, has recommended that RSNA continue to support QIBA and listed specific recommendations about such topics as assuring compliance and applicability to clinical care.

b. *In partnership with NCRR/NIH, RSNA provides support for a CTSA Imaging Working Group which constitutes another component of a collaborative, multidisciplinary infrastructure to foster research, approval and use of quantitative imaging biomarkers.*

The CTSA Imaging Working Group held a teleconference every other month to facilitate communication and sharing of best practices among funded CTSA sites on issues relevant to imaging in clinical research. On alternate months, the CTSA Steering Committee met. In-person educational sessions were held by CTSA at the SCTS meeting in April 2011 and at the in-person ACRIN meeting in September 2011.

Leadership of and responsibility for the CTSA Imaging Working Group has been transferred to the RSNA Research Development Committee (RDC), in part because of the NIH reorganization involving NCRR and resulting uncertainties around the CTSA program. The RDC has created a Vice Chairs of Radiology committee, and that committee has created three sub-groups that are now addressing the issues initiated within the CTSA IWG.

c. *We have created an Ad Hoc Committee on Open Image Archives which will provide in approximately 6 months a report containing recommendations for creating one or more national repositories of quantitative imaging biomarker data.*

This committee has continued to meet and has produced proposals presently being considered by the Steering Committee. The task force has delivered several 'use cases' as well as other documents supportive of defining proposals for implementation. The steering committee decided to form a joint QIBA/Radiology Informatics Committee which met in a face-to-face session in September 2011 to consider how to best move this agenda forward. An action plan is being followed to converge on a specific proposal to the RSNA Board in the March 2012 time-frame.

The *ad hoc* Open Image Archives (OIA) Committee of QIBA was formed to assess what could be done to improve the creation and sustained growth of imaging archives. These efforts have transitioned to a committee of combined QIBA, OIA and Radiology Informatics Committee (RIC) members which was formed to provide a platform of collaboration, to optimize synergy between overlapping areas of expertise and interests, and, in particular, to support and augment the OIA activities by drafting a plan for potential RSNA involvement for imaging data warehouses going forward.

Four classes of QIBA use cases were defined: A. Comparative Evaluation of Imaging Biomarker Performance versus Gold Standard; B. Public Resource Shared Data (*e.g*., Image Processing Algorithm Development); C. FDA Approval of Clearance of Imaging Tests; and D. Pharma Clinical Trials with Imaging Biomarkers as Endpoints; for two quantitative imaging

biomarker projects: 1) CT volumetric image analysis for management of patients with lung cancer, and 2) quantification of tumor metabolism using FDG-PET standardized uptake value (SUV) image analysis.

Imaging data warehouse needs for each of the QIBA Technical Committee Working Groups (DCE-MRI, FDG-PET, Volumetric-CT, fMRI, and COPD-Asthma) were summarized and common features noted. These included the requirement to accommodate different image and non-image data formats (including, in addition to DICOM image files, a variety of other file formats such as XML, TIFF, NiFTI, etc.) and a wide variety of relevant clinical metadata. In addition, the following needs were identified: data input and search and query-retrieve capabilities; image de-identification, data security and user authentication with group sharing; and data output statistics and analytics functions, though not necessarily image display applications.

Existing tools and databases, including The Cancer Imaging Archive (TCIA), the National Biomedical Image Archive (NBIA), Laboratory of NeuroImaging (LONI), eXtensible Neuroimaging Archive Toolkit (XNAT), and MIDAS, were examined, and current limitations detailed. Of concern is the lack of and/or need for a "trusted third party", the need to promote a culture of sharing perhaps with a reward system or participation, and a business model for long-term sustainability. Additional limitations of existing image data archives include ease-of-use regarding tool downloads, data uploading, tool configurability and functional enhancements. The need for front-end image and metadata collection tools, security control, advanced search, and back-end data analytics components was also noted.

The combined QIBA, OIA and Radiology Informatics Committee (RIC) members have produced a proof-of-concept implementation using a pilot project from the MRI modality committee and have begun to load additional data. The joint QIBA/RIC committee received approval to deploy this seed project for an image warehouse with associated capability from the RSNA Board. The "Quantitative Imaging Data Warehouse" (QIDW) is currently in pilot phase.

    d. *RSNA staff supported by this NIBIB contract will provide project management and staff support for same.*

Staff has successfully met the challenges as well as the opportunities afforded by this contract assignment.

NIBIB Task 19. Explore self-funding models to maintain forward progress of the infrastructure and effort described in task 18 above.

We will create an Ad Hoc Task Group to conduct strategy discussions on this topic during Year 1 and will develop a draft proposal by year end. Based on the nature of that proposal we will lay out actions and a plan for Year 2.

Pursuant to the activity described in Task 18a, an ad hoc task force has been set up to develop ideas which have been considered for creating a self-funding program. The QIBA Task Force on Commercial Business Models held a face- to- face meeting to develop ideas stemming from a variety of sources. This work is ongoing.

B. FUNDED PROJECTS

A listing of funded and completed projects is provided below. Summary information is included in Sections C-E, and full text reports are provided as attachments to this report.

CT, round 1:

Biomarker	Title	Amount Awarded	Submitter	TimeLine (contract)	Project Description (Abstract)
VolCT	Inter-scanner/Inter-clinic Comparison of Reader Nodule Sizing in CT Imaging of a Phantom	$14,000	Michael McNitt-Gray, PhD (UCLA)	04/01/2011 - 03/31/2012	Project Description Inter-scanner/inter-clinic comparison of reader nodule sizing in CT imaging of a phantom.
VolCT	Inter-scanner/Inter-clinic Comparison of Reader Nodule Sizing in CT Imaging of a Phantom	$11,000	David Clunie, MBBS (CoreLab Partners)	04/01/2011 - 03/31/2012	Project Description Inter-scanner/inter-clinic comparison of reader nodule sizing in CT imaging of a phantom (reader mark-up services).
VolCT	Assessing Measurement Variability of Lung Lesions in Patient Data Sets	$13,185	Michael McNitt-Gray, PhD (UCLA)	04/01/2011 - 03/31/2012	Project Description The purpose of this project is to perform the statistical analysis of data collected under QIBA Volumetric CT committee's 1B experiment, which is investigating the minimum detectable
VolCT	Validation of Volumetric CT as a Biomarker for Predicting Patient Survival	$62,495	Binsheng Zhao, DSc (Columbia Univ)	04/01/2011 - 12/31/2012	Project Description Unidimensional measurements have become a de-facto standard for assessing a patient's response to therapy. In essence, the RECIST measurement is a "surrogate" for tumor burden and change in this metric is used to guide drug discovery. We and other groups have demonstrated that the unidimensional measurement and change in the unidimensional measurement do not always correlate with change in tumor burden. The actual change in tumor burden is better assessed by change in tumor volumes as measured on CT. We plan to retrospectively analyze tumor burden change in patients enrolled on an already completed large, multicenter Phase II/III clinical trial in metastatic colorectal cancer. We are in an excellent position to validate CT volumetric response assessment technique using our computer algorithms and the clinical data in this trial. 451 patients have been enrolled in this trial and have measurable target lesions in the liver as well as in the lungs and lymph nodes. Each patient underwent an average of 5 CT scans (baseline and follow-ups at every 6-week after the treatment and every 3-month starting 49 months until disease progressed or patient died). The trial also collected a number of tissue biomarkers (e.g., markers in the IGF-1R or EGFR signaling pathway) before and at 6-week after the treatment and tumor genotype from blood sample taken pre- and every 6-week post-treatment. Volumetric (as well as unidimensional and bidimensional) measurements of target lesions on all scans of all patients will be calculated using our in-house computer-aided (CA) methods developed for segmentation of lung, liver and lymph node metastases. Intra- and inter-reader variability of the three measurements will be explored in a subset of patients (i.e., the first 50 patients who have more than 3 scans). Finally, correlations of tumor response and time-to-progression assessed unidimensionally, bidimensionally and volumetrically with clinical outcome (overall survival), tissue biomarkers and tumor genotype will be performed. By evaluating the magnitude of measurement variability and the optimal correlation, cut-off value or continuous change variables to identify tumor regression and progression can be established for unidimensional (ie. modified RECIST), bidimensional (ie. modified WHO) and volumetric response assessment methods. If proven successful and accepted by the oncology community and regulatory agencies, this research will aid the discoveries of cancer drugs and tissue biomarkers as well.
VolCT	Development of Assessment and Predictive Metrics for Quantitative Imaging in Chest CT	$50,000	Ehsan Samei, PhD (Duke)	04/01/2011 - 09/29/2012	Project Description: Previous work in our laboratory has developed a framework for predicting quantitative imaging performance from basic system performance measurements.1 These figure of merits (FOM) included metrics that characterize the resolution (modulation transfer function, MTF) and noise (noise power spectrum, NPS) of the image. It was shown that the precision with which medical images can be used to estimate volume of lesions can be predicted from these simpler FOM. By extending this framework to CT, we expect to be able to evaluate how well clinical CT systems perform various quantitative imaging tasks based on measurements of system FOMs (i.e., noise, contrast, and spatial resolution) via conventional QA phantoms. The goal of this project is to utilize these models in the evaluation of existing phantoms in the characterization of quantitative performance of CT and further develop a calibration procedure to assess compliance of quantitative imaging technique in volumetric CT. As part of this project, we anticipate the collaboration with existing QIBA efforts (1a, 1b, and 1c).
VolCT	Measurement of Pulmonary Nodule (Solid, Part-Solid and Ground Glass) Volume, Longest Diameter and CT Attenuation Resulting from Differences in Reconstruction Thickness, Reconstruction Plane, and Reconstruction Algorithm	$42,070	Kavita Garg, MD (U Colorado)	04/01/2011 - 09/29/2012	Recently released initial results of the National Lung Screening Trial (NLST) show mortality reduction by 20% in the CT arm compared with CXR. If screening becomes widely adopted in those at high risk, follow-up investigation of positive scans will impose a major burden on the health care system. In patients with positive scans, a risk stratification strategy or quantitative analysis of lung nodules could reduce this burden by reducing the rate of follow-up in those who are determined to be at lower risk. Quantitative CT analysis for solid nodules has been attempted previously, however there is no significant data available for subsolid nodules.
	TOTAL CT FUNDING AWARDED	**$192,750**			

CT, round 2:

Biomarker	Title	Budget request	Submitter	TimeLine (contract)	Project Description (Abstract)
VolCT	Validation of Volumetric CT as a Biomarker for Predicting Patient Survival (Carry-over from Round-1)	$62,495	Binsheng Zhao, DSc (Columbia Univ)	04/01/2011 - 12/31/2012	Project Description The purpose of this project is to extend the data collection and statistical analysis of the QIBA Volumetric CT committee's 1B experiment, which is investigating the minimum detectable change in lesion size from patient datasets imaged on CT. That project used: (a) "Coffee Break" CT image datasets from 32 NSCLC patients who were imaged twice over a short (15 minute) interval on the same scanner using thin (1.25 mm) slices; (b) one lesion was identified for each patient, (c) Image data was marked up by five radiologists at RadPharm (now CoreLabs); (d) each reader marked the lesions on each of the repeat scans to obtain measures of volume, single longest diameter and bi-dimensional diameters. This data was previously collected and initial analyses have been performed.
VolCT	Development of Assessment and Predictive Metrics for Quantitative Imaging in Chest CT (Carry-over from Round -1)	$25,000	Ehsan Samei, PhD (Duke)	04/01/2011 - 09/29/2012	Project Description The purpose of this project is to extend the data collection and statistical analysis of the QIBA Volumetric CT committee's 1B experiment, which is investigating the minimum detectable change in lesion size from patient datasets imaged on CT. That project used: (a) "Coffee Break" CT image datasets from 32 NSCLC patients who were imaged twice over a short (15 minute) interval on the same scanner using thin (1.25 mm) slices; (b) one lesion was identified for each patient, (c) Image data was marked up by five radiologists at RadPharm (now CoreLabs); (d) each reader marked the lesions on each of the repeat scans to obtain measures of volume, single longest diameter and bi-dimensional diameters. This data was previously collected and initial analyses have been performed.
VolCT	Extension of Assessing Measurement Variability of Lung Lesions in Patient Data Sets: Variability Under Clinical Workflow Conditions	$14,110	Michael McNitt-Gray, PhD (UCLA)	08/01/2011 - 09/29/2012	Project Description QIBA 1A study investigated the bias and variance using CT images of an anthropomorphic phantom and obtained measurements using only one algorithm. However, image processing algorithms have been developed in many organized activities from a number of groups using many different approaches with varying amounts of human interaction and different levels of segmentation success. This study proposes to investigate the effects of different algorithms in bias and variance using reference data sets of both phantoms and patients; in addition, this study proposes to investigate the performance of different methods and degree of automation in the algorithm. This study can increase knowledge for the QIBA Profile and to provide a context in which multiple parties have incentives to participate. (1) manual analysis in which sites perform required measurement task and manually supply data back to QIBA 3A project personnel and (2) scripted analysis that would apply to multiple data sets to invest in a re-usable infrastructure for large-scale algorithm testing
VolCT	Extension of Assessing Measurement Variability of Lung Lesions in Patient Data Sets: Variability Under Clinical Workflow Conditions (reader mark-up)	$13,125	David Clunie, MBBS (CoreLab Partners)	08/01/2011 - 07/31/2012	
VolCT	Comparative Study of Algorithms for the Measurement of the Volume of Lung Lesions: Assessing the Effects of Software Algorithms on Measurement Variability	$35,500	Michael McNitt-Gray, PhD (UCLA)	08/01/2011 - 07/31/2012	
COPD	Impact of Dose Saving Protocols on Quantitative CT Biomarkers of COPD and Asthma	$49,754	Sean Fain, PhD (Univ of Wisconsin)	08/01/2011 - 12/31/2012	Longitudinal clinical research studies that depend on quantitative measures of CT (qCT) number and structures depend on stable and reproducible image fidelity in often young, radiation sensitive, populations. Quantitative image measures of CT number and structural dimensions in ongoing lung studies, such as parenchymal density and airway wall thickness, have been shown to vary substantially according to the CT scanner make and model due to factors that differentially affect image noise, artifacts, and other image quality parameters, such as low contrast detectability and MTF. Imaging at the lowest possible dose, while maintaining diagnostic accuracy, is important to reduce the patient's risks from radiation. While image noise is normally adversely affected by adjusting scan parameters to decrease radiation dose, there now exist methods that allow decreased dose with no significant effect on image quality. However, there are significant questions as to effects of these techniques on qCT measures. These dose reduction techniques include automatic exposure control (AEC) systems, such as Smart mA on the GE systems, which produce mA modulation in the Z and angular directions, and iterative reconstruction (IR) techniques such as ASIR (GE).
	TOTAL CT FUNDING AWARDED	$199,984			

MR, round 1:

DCE-MRI	DCE-MRI Phantom Fabrication, Data Acquisition and Analysis, and Data Distribution	$60,347	Edward Jackson, PhD (MD Anderson CC)	04/01/2011 - 12/31/2012	Project Description The DCE-MRI Technical Committee has developed a prototype DCE-MRI phantom that is proposed for use in DCE-MRI clinical trial site qualification as well as ongoing quality control processes. When it was initially formed, the DCE-MRI Technical Committee identified the need for such a robust DCE-MRI contrast response phantom as a top priority. Since that time, the committee developed a "generic" DCE-MRI acquisition protocol, developed a protocol for quality control / initial site qualification, each process a key component of the first DCE-MRI Profile, and has evaluated the use of a modified ADNI MagPhan phantom and, upon initial multicenter testing of two copies of this phantom design, found the phantom to be insufficiently robust to shipping and handling, too limited in its evaluation of R1 (longitudinal relaxation rate) contrast response assessment, and not time efficient in routine application. A new phantom design was proposed to address each identified weakness of the modified ADNI MagPhan phantom for DCE-MRI contrast response characterization, relaxometry, and quality control.
DCE-MRI	Software Development for Analysis of QIBA DCE-MRI Phantom Data	$29,975	Edward Ashton, PhD (VirtualScopics)	04/01/2011 - 03/31/2012	Project Description This project will address the development of a distributable software package to allow the analysis of QIBA DCE-MRI phantom data.
DCE-MRI	Digital Reference Object for DCE-MRI Analysis Software Verification	$57,763	Daniel Barboriak, MD (Duke)	04/01/2011 - 09/29/2012	Project Description One barrier to implementation of dynamic contrast-enhanced (DCE) MRI in multi-center clinical trials is that available software packages used to analyze the images may differ in their approach and implementation, causing variability in the extracted quantitative parameters. Because no standardized image analysis method is available, results obtained using DCE-MRI in different laboratories are difficult to compare, and the rational choice of one software implementation over any other for use in a multi-center trial is exceedingly challenging. As a first step in providing a standardized analysis process, it is necessary to ensure that software implementations are extracting parameters accurately. In this project, we propose to create digital reference objects (DROs) using synthetic data in order to help verify software packages for use in DCE-MRI analysis, and to initiate the development of verification protocols as a method to qualify software packages for use in clinical trials of DCE-MRI.
fMRI	Quantitative Measures of fMRI Reproducibility for Pre-Surgical Planning	$19,411	Edgar DeYoe, PhD (Med College of Wisconsin)	04/01/2011 - 08/31/2012	Project Description This subproject will provide quantitative measures of reproducibility for a unique set of vision- and motor-related fMRI brain maps using a set of data manipulations, computations, and the AMPLE normalization algorithm that will be standardized in coordination with subproject 1. In addition, this subproject will provide a unique analysis of the reproducibility of fMRI functional specificity and will compare reproducibility measures over different time durations both within and across subjects. The results of this study will help address NIBIB Tasks 1, 3, 6, 7, and 9 in the context of fMRI as a biomarker of brain function/dysfunction.
fMRI	Quantitative Measures of fMRI Reproducibility for Pre-Surgical Planning-Development of Reproducibility Metrics	$33,423	James Voyvodic, PhD (Duke)	04/01/2011 - 07/31/2012	Project Description This sub-project will develop metrics for quantifying reproducibility within and across fMRI scans, and it will apply those metrics to existing data sets to assess the reproducibility of fMRI results both within and across scanning sessions. The metrics to be developed will be based on using the AMPLE normalization algorithm (Voyvodic, 2006) to assess reproducibility and will include both voxel-wise and ROI-based measures of the consistency of fMRI activation maps over time. The results of this study will help address NIBIB Tasks 1, 3, 6, 7, and 9. Demonstration and quantification of reproducibility is an essential step in the development of the QIBA Profile for fMRI.
	TOTAL MR FUNDING AWARDED	**$200,919**			

MR, round 2:

DCE-MRI	Test-Retest Evaluation of Repeatability of DCE-MRI and DWI in Human Subjects	$175,000	Mark Rosen, MD, PhD (ACR/U Penn)	08/01/2011 - 02/28/2013	**Project Description** In order to evaluate the profile claims for the repeatability of tumor DCE-MRI metrics, it is ultimately necessary to evaluate the performance of the QIBA DCE-MRI protocol in human subjects with tumors. This process would include not only implementing the imaging protocol, but also introducing site imaging personnel to proper procedures for magnet qualification, the selection of tumor-bearing human subjects, and the implementation of the proper DCE-MRI imaging exam to these subjects. Through discussions within the DCE-MRI subcommittee, it has been decided to seek a partnership with the American College of Radiology Imaging Network (ACRIN) to facilitate the development and ultimately the implementation of the test-retest DCE-MRI human study. The overall project, including the completion of the protocol, selection of participating sites, central and site regulatory approval, and implementation is expected to last two years.
fMRI	Validation of Breath Hold Task for Assessment of Cerebrovascular Responsiveness and Calibration of Language Activation Maps to Optimize Reproducibility	$29,376	Jay Pillai, MD (Johns Hopkins MC)	08/01/2011 - 09/29/2012	**Project Description** Assessment of the integrity of cerebral vascular responsiveness (CVR) is a critical problem in clinical applications of fMRI brain mapping, and must be addressed in any attempt to quantitate fMRI results in patients. Neurovascular uncoupling (NVU) associated with brain tumors and other brain diseases may result in regional variations in CVR that affect the ability to generate a BOLD signal and thereby reliably and reproducibly localize eloquent cortex during presurgical mapping. Multiple studies have demonstrated that a breath-hold (BH) hypercapnia task is a reliable tool for assessing CVR and normalizing BOLD response among different subjects, different brain regions and various features of the scanning environment. In this project we plan to 1) validate use of a BH task for mapping of brain CVR and 2) use such maps to calibrate language task-based BOLD activation maps in order to both reduce intersubject variability and increase intrasubject reproducibility across scan sessions. For the first objective, we will compare BH CVR maps with T2* DSC MR perfusion imaging maps using quantitative region of interest analysis to assess concordance of regions of decreased CVR with regions of abnormal perfusion in a cohort of 10 brain tumor patients. For the second objective, we plan to apply the normalization/calibration technique described by Thomason et al. (2007) to an existing dataset of approximately 10 normal right-handed native English speaking subjects who performed two BOLD language tasks--silent word generation and sentence completion--in addition to a breath hold (BH) task, as well as to a cohort of a similar number of brain tumor patients who performed similar tasks. Furthermore, for the patient cohort, analysis of multiple runs of the same language activation paradigms will assess intrasubject BOLD activation variability utilizing CVR-calibrated activation maps. The results of this study will help fill the high priority gaps of evaluation of neurovascular responsiveness, reproducibility and protocol optimization, defined by the QIBA fMRI subcommittee, and address NIBIB Tasks 1, 3, 6, 7, and 9 in the context of fMRI as a biomarker of brain function/dysfunction.
	TOTAL MR FUNDING AWARDED	**$204,376**			

NM, round 1:

FDG-PET/CT	Meta-analysis to Analyze the Robustness of FDG SUV Changes as a Response Marker, Post and During Systemic and Multimodality Therapy, for Various Types of Solid Extracerebral Tumors	$73,000	Otto Hoekstra, MD (VU Med Ctr, NL)	04/01/2011 - 03/31/2012	Project Description Meta-analysis on the association between FDG uptake changes and histopathological response. Validated qualitative and quantitative statistical model of quantitative FDG PET to predict response of neoadjuvant therapy in solid tumors.
FDG-PET/CT	QIBA FDG-PET/CT Digital Reference Object Project	$68,240	Paul Kinahan, PhD (U Washington)	04/01/2011 - 03/31/2012	Project Description Construct a common reference DICON PET/CT test image (Digital Reference Object or DRO) as if generated by each vendor's PET/CT scanner. This will then be read on PET/CT display stations to check SUV computation fidelity and region of interest (ROI) analysis performance. This is motivated by the vendor-specific variations in the standardized uptake value (SUV) calculations. It is well known that variations in the implementation of DICOM standards produce substantial quantitative differences in SUVs for the same image on different display stations. In addition, the behavior of ROI analysis tools, (e.g., due to image interpolation) is rarely, if ever, quantitatively understood.
FDG-PET/CT	Analysis of SARC 11 Trial PET Data by PERCIST with Linkage to Clinical Outcomes	$57,500	Richard Wahl, MD (Johns Hopkins MC)	04/01/2011 - 02/28/2013	Project Description The insulin-like growth factor (IGF) pathway plays an important role in a variety of physiological processes in humans and animals, including normal growth and development. Additionally, this pathway has been shown to play an important role in the development of conditions like cancer. IGF signaling has been proposed to play a major role in the very aggressive nature of certain sarcomas, like Ewing's sarcoma family of tumors (ESFT) and synovial sarcomas The SARC 11 trial prospectively evaluated the utility of an anti Insulin like Growth Factor Human Monoclonal Antibody (R1507) as monotherapy for Sarcomas of several types (Recurrent or Refractory Ewing's Sarcoma, Osteosarcoma, Synovial Sarcoma, Rhabdomyosarcoma and Other Sarcomas.) . All patients received R1507 9mg/kg i.v. . This single arm study evaluated the efficacy and safety of R1507 in patients with recurrent or refractory sarcoma Clinical efficacy of the trials was judged by: Objective response rate [Time Frame: Week 24, and every 12 weeks thereafter] , Progression-free survival in patients with Ewing's sarcoma [Time Frame: Week 18] , Duration of response, PFS, and overall survival. In this multicenter study, a baseline PET scan was obtained as well as a PET scan at approximately 9 days post therapy initiation. A follow up scan at 12-18 weeks was obtained in those remaining on study. A total of 311 patients entered the study. Importantly, the PET was not used to alter the therapy. Thus, the PET data can be used to determine prognostic ability. Accrual to the study is now complete and the full PET data set have been collected and are available in DICOM form for analysis. Details of the SARC 11 trial and the rationale for anti IGF antibody therapy are detailed at Clinical Trials. gov and the SARC website. http://www.clinicaltrials.gov/ct2/show/related/NCT00642941?term=r1507 and http://www.sarctrials.org/SARC011r1507 We propose to analyze the PET data quantitatively (and qualitatively) using PERCIST and EORTC response criteria to determine how predictive PET, notably changes in PET signal between baseline and the first follow up scan, is of clinical outcomes. We will also examine inter-observer consistency. We will apply commercial software and our in house developed software for analysis. This trial is particularly suitable for analysis of PET data as they are not
	TOTAL NM FUNDS AWARDED	**$198,740**			

NM, round 2:

FDG-PET/CT	Personnel Support for FDG-PET Profile Completion	$16,000	Eric Perlman, MD (Perlman Advisory Group)	10/01/2011 - 12/31/2011	**Project Description** This is a proposal for expert support as a 'Profile Writer' to convert the considerable amount of material accumulated by the FDG-PET TC into a QIBA Profile.
FDG-PET/CT	Evaluation of the Variability in Determination of Quantitative PET Parameters of Treatment Response across Performance Sites and Readers	$100,000	Richard Wahl, MD (Johns Hopkins MC)	10/01/2011 - 02/28/2013	**Project Description** There is very limited data on the performance of varying readers and quantitative imaging workstations in determining cancer treatment response using FDG PET/CT. We propose a study design using well-defined anonymized pre-treatment and post-treatment FDG PET scans of cancer patients as an analysis set. All
FDG-PET/CT	PERCIST Validation	$50,000	Otto Hoekstra, MD (VU Med Ctr, NL)	04/01/2012 - 12/31/2012	**Project Description** Validate PERCIST metrics using peer-reviewed patient cohorts.
FDG-PET/CT	Evaluation of FDG-PET SUV Covariates, Metrics, and Response Criteria	$34,000	Jeffrey Yap, PhD (Dana-Farber Cancer Institute)	02/01/2012 - 12/31/2012	**Project Description** We have developed a large database of more than 25,000 PET oncology studies, which includes critical acquisition parameters, patient information, and DICOM CT and PET images.1 Many of these studies are from multi-center trials that included PET scanner qualification, phantom imaging, central review and PET SUV analysis, and collection of clinical outcome data. We propose to perform a retrospective meta-analysis to compare different PET metrics, response assessment criteria (EORTC, PERCIST), PET SUV covariates (FDG dose, glucose, fasting time, patient size, etc.), and clinical outcome. A small component of this activity has already been performed in a subset of data comparing the impact of metabolic response assessment using SUVmax vs. SUVmean (Figure 1) and SUV patient size normalization using lean body mass vs. body weight (Figure 2).2,3,4 The requested resources that are needed to complete this work include the compilation of images, meta-data, and clinical trial outcome measures from a research miniPACS archive, multiple clinical trial MS Access databases, a clinical PET database, and various sources of clinical trials results such as Excel spreadsheets. In addition to the existing results, additional image analyses will be performed to generate normal tissue ROIs (e.g. liver) as well as multiple tumor ROIs for studies that only included single tumor per patient in the original analysis. A software package will be developed in IDL to establish a DICOM server research archive and automatically extract and compare various PET metrics (e.g. SUVmax, SUVmean, SUVlbm) from previously performed ROI analysis. This will address a major limitation in commercial software that only allows the use of a single metric and/or response criteria for a given study and facilitate the automated generation and comparison of different PET metrics and response criteria. Lastly statistical analysis will be performed on the results of multiple clinical trials in order to evaluate the impact of covariates, PET metrics, and response criteria on the performance of FDG-PET SUV as an imaging biomarker of therapeutic response. This will yield critical results for supporting claims in the QIBA profile with such as the variability in response assessment using different methods as well as justify consensus recommendations in the UPICT protocol, e.g. with regards to image analysis and response assessment.
	TOTAL NM FUNDING AWARDED	**$200,000**			

Cross-modality:

Cross Modality	Groundwork for QIBA Image Reference Database	$10,250	Gudrun Zahlmann, PhD (Roche)	04/01/2011 - 03/31/2012	**Project Description** QIBA profiles of all technical committees describe what the imaging process for a certain quantitative imaging biomarker should look like in order to fulfill the profile claims. The claims state how accurate and reliable the respective quantitative imaging biomarker can be expected if the profile is followed. The so-called bull's-eye principle is used for describing ideal, target and acceptable quality. All teams deal with phantom imaging to assess the scanner quality. In addition clinical images are captured and analyzed in order to define the claims. It would be beneficial to support the QIBA profiles with a repository of phantom and clinical images illustrating the QIBA understanding of ideal, target and acceptable quality in real life image examples.
Cross Modality	Groundwork for QIBA Image Reference Database	$16,000	Rick Avila, MS (Kitware)	04/01/2011 - 10/31/2012	**Project Description** QIBA profiles of all technical committees describe what the imaging process for a certain quantitative imaging biomarker should look like in order to fulfill the profile claims. The claims state how accurate and reliable the respective quantitative imaging biomarker can be expected if the profile is followed. The so-called bull's-eye principle is used for describing ideal, target and acceptable quality. All teams deal with phantom imaging to assess the scanner quality. In addition clinical images are captured and analyzed in order to define the claims. It would be beneficial to support the QIBA profiles with a repository of phantom and clinical images illustrating the QIBA understanding of ideal, target and acceptable quality in real life image examples.
	TOTAL CROSS FUNDS AWARDED	**$26,250**			

C. QUANTITATIVE CT

Completed during the contract term:

- With respect to CT Volumetry:
  - Experimental groundwork:
    - ***Inter-scanner/Inter-clinic Comparison of Reader Nodule Sizing in CT Imaging of a Phantom*** (Fenimore): Project complete. <u>Conclusions</u>: The Figure shows that for nearly all scanners, the absolute percent relative error was well below 15% for spherical lesions that are ≥ 10 mm in diameter, regardless of scanner or protocol used; for spiculated lesions, the

absolute percent errors were higher, but only a few exceeded 15% (5 out of 154 and most of those were confined to one scanner, regardless of protocol). In the spiculated lesions, 10 out of 154 were greater than 15%.

- ***Validation of Volumetric CT as a Biomarker for Predicting Patient Survival*** (Zhao): Project complete. Conclusions, Part 1. Measurement Variability. Completed the intra- and inter-reader variability study and successfully delivered the proposed report. Since variability in interpretation of change in total tumor burden can occur during both selection and measurement of target lesions, study was designed to consider both of these factors.

  Conclusions, Part 2. Clinical Correlations. After locking longitudinal tumor measurements, project was provided the patient survival data and further clinical information by the company. Out of the 560 patients, 478 patients (147 died and 331 alive) have a baseline and at least one follow-up scan and are qualified for further analysis in terms of the comprehensiveness of imaging data. Since the date of the last follow-up for those alive is not available, patients were censored at the time of their last scan or on the date of discontinuation from the study, whichever was later. The company that had provided the data for this study is retrieving the missing survival data for us now. For this reason, we only analyzed the unidimensional data. However, the methods can be applied directly to analyze bidimensional and volume data once we receive the missing clinical information. Our preliminary results indicate that the minor change category of -30% -10% may correlate with longer survival compared to the -10% - 20% group.

- ***Development of Assessment and Predictive Metrics for Quantitative Imaging in Chest CT*** (Samei): Project complete. Conclusions: Publications: Chen B, Richard S, Barnhart H, Colsher J, Amurao M, Samei E. Quantitative CT: technique dependency of volume assessment for pulmonary nodules. Physics in Medicine and Biology 57: 1335–1348, 2012; and Chen B, Richard S, Samei E. Relevance of MTF and NPS in quantitative CT: towards developing a predictable model of quantitative performance. SPIE International Symposium on Medical Imaging, San Diego, CA, February 2012, Proc. SPIE Medical Imaging, 2012.

- ***Quantifying Variability in Measurement of Pulmonary Nodule (Solid, Part-Solid and Ground Glass) Volume, Longest Diameter and CT Attenuation Resulting from Differences in Reconstruction Thickness, Reconstruction Plane, and Reconstruction Algorithm*** (Garg): Project complete. Conclusions; Hypothesis One: The absolute percent error in part-solid nodule tumor measurements will be less than or equal to 15% (the QIBA Solid Nodule > 10mm claim). In this study, the absolute percent error for part-solid nodule measurements was found to be 20.38 (17.75), for nodules of diameter 10 and 20mm. This value was significantly dependent on the reader, nodule diameter, nodule outer shape and nodule core density. Further work is needed to investigate these dependencies for part solid measures. It was also noted that the readers consistently obtained volumes that were larger than the actual volume. Further study is needed on the proper display for visualizing the PS lesion with respect to the displayed segmentation outline. Concerns have been raised regarding the actual "ground truth" values of the nodules; this will be investigated further. Absolute bias measures for solid lesions were found to be significantly worse than those of the part-solid lesions. In addition to lesion size differences, it would be important to see if the algorithm designed for part-solid lesions performs less effectively on solid lesions. Finally, the one-dimensional measure, SA-RECIST exhibited significantly less absolute error than the volume measure. Noting that the 1-D measure results from the same segmentation as the volume measure and for a spherical lesion, the volume goes up as radius-cubed, one might actually expect that the volume measures should be worse than they were. For this study, all of the cores and 50% of the part-solid lesions were spherical. There was also a

significant dependence on nodule outer shape; further study is required to understand this trend. Hypothesis Two: Dose will not significantly affect the part-solid absolute percent error measurement for the nodules chosen for this study. This study confirmed a lack of significant dependence of any morphometric measure on dose. Hypothesis Three: Part-solid nodules will exhibit increased inter-reader and intra-reader measurement variability, compared to solid nodule studies. This study confirmed greater reader variability for all nodules measured and a poor ICC for measuring small changes in lesion volume. Further investigation into the user interface and process are needed to determine the reason for this.

- ***Assessing Measurement Variability of Lung Lesions in Patient Data Sets: Variability under Clinical Workflow Conditions*** (McNitt-Gray and Clunie): Project complete. Discussion and Conclusion: These results demonstrate that the measurement variability has decreased markedly when using the sequential locked reading paradigm of this study when compared to the independent reading paradigm of the previous study. The decrease in measurement variability was seen for all lesions and for all measurements, but the reduction in variability appears to have been greatest in volumetric measurements. The results from this study should be directly applicable to the QIBA profile and its descriptions regarding "best practices" for clinical trials and the reduction of measurement variability. Specifically, the use of a sequential locked reading paradigm rather than independent randomized reading may allow for a smaller biological change to be detected in an individual subject. Further analyses yet to be performed include the reduction in measurement variability of lung lesions that meet the QIBA profile criteria (i.e. that a "given tumor is measureable [tumor margins are sufficiently conspicuous and geometrically simple enough to be recognized on all images in both scans] and the longest in-plane diameter of the tumor is 10 mm or greater"), as well as clear determination of the statistical significance of the difference in reading paradigms.

- ***Comparative Study of Algorithms for the Measurement of the Volume of Lung Lesions*** (Athelogou PI, statistical analysis Kim): Project complete. Conclusion: The performance of analysis methods were highly differentiated, but the majority fell within 15% absolute percent error across participants. These results match with QIBA performance and target levels, with providing insight for variation of software programs in measuring volumes. Next Steps of the QIBA 3AGroup: Paper preparation for pilot and pivotal study and preparation of the next challenge. For the next challenge clinical data will be used (QI-Bench datasets).

  o Profiling:
      ▪ The team conducted a public comment period, addressed all comments, and released its advanced disease profile in August 2012.
      ▪ A new writing group has started, covering Lung Nodule Assessment in CT Screening.
- The COPD/Asthma committee has:
  o Characterized various foam inserts and other aspects of phantom design for effective calibration and quality control in lung densitometry studies.
  o ***Impact of Dose Saving Protocols on Quantitative CT Biomarkers of COPD and Asthma*** (Fain): Project complete. Accomplishments: 1. We have established the impact of dose on quantitative measures of wall thickness, and microstructural density using the COPD gene CT phantom. 2. We have also established the impact of a product non-linear iterative reconstruction (ASIR) designed to preserve image quality under low x-ray dose image acquisition conditions on quantitative measures of wall thickness, and microstructural density using the COPD gene CT phantom for low and high radiation dose scans. 3. We have also extended the results of

measurements in phantoms established in 1 and 2 above to human lung images utilizing raw data reconstructions of human lung studies. We have been able to reproduce results for sampling density with a reduced display FOV reconstruction to improve measures of wall thickness; demonstrate the impact of measurement kernel on wall measures using Airway Inspector vs. VIDA; and have determined the consequences of nonlinear reconstruction (ASIR) on quantitative airway and microstructural lung parenchymal density measures. 4. We have also established the impact of a non-linear adaptive statistical iterative reconstruction (ASIR) that under certain acquisition and reconstruction settings can preserve image quality under low x-ray dose conditions and support accurate airway morphology measures. This project is the first to explore the consequences of ASIR on quantitative measures of wall thickness, and microstructural density using the COPD gene CT phantom for low and high radiation dose scans. This includes numerical determination of the limits of spatial resolution and impact of reconstruction kernel on accuracy of wall measures as a function of radiation dose when using ASIR reconstruction. This component of the project has established numerical improvements in airway morphology measures that are maintained at low X-ray by using the advances from point 2 above in combination with the ASIR reconstruction. 5. We have extended the results of measurements in phantoms established in 2 and 4 above to measures of human airways and lung tissue density by retrospective reconstruction of raw data archived from human lung studies. We have been able to show improved depiction of airway walls and maintenance of contrast and quantitative measures with ASIR. These results have given us confidence that we are able to reproduce improvements observed in the COPD Gene phantom to human subjects. We have performed these reconstructions in only 4 raw data sets and intend to expand this to ~12 raw data sets to establish consistency of the resolution optimized protocol. The final results have extended the range of parameters identified using the phantom experiments to human subjects by retrospective reconstruction of raw projection data. These reconstructions explore spatial resolution and noise characteristics of CT reconstruction with high bandwidth kernels with and without ASIR to improve accuracy of airway measurements while mitigating noise amplification. Results point to a novel reconstruction strategy/parameter space that can improve airway measures for IR approaches.

- o First draft Profile is in progress.

D. QUANTITATIVE FDG-PET

Completed during the contract term:

- Experimental Groundwork:

  - o ***Meta-analysis to Analyze the Robustness of FDG SUV Changes as a Response Marker, Post and During Systemic and Multimodality Therapy, for Various Types of Solid Extracerebral Tumors*** (Hoekstra): Project Complete. Results: sixty-nine studies were identified reporting on 88 data sets (20 during-treatment PET assessments, 68 end-of-treatment PET assessments). QUADAS revealed heterogeneity but it was not useful to stratify studies according to quality aspects relevant for the study question. Of these 36 (52%) reported individual level data and 44 (64%) used SUVmax as the measure of uptake. The most common tumor types were esophagus 18 (26%), rectal 15 (22%), sarcoma 14 (20%) and breast (14%), with 30 (43%) studies reporting chemotherapy (CT) treated, 36 (52%) reporting chemoradiotherapy (CRT), and 3 (4%) reporting combined CT /CRT cohorts. A heterogeneity test indicated excessive variation in the ratio of PET values (post-therapy/baseline) for responders and non-responders in breast cancer cohorts, potentially due to differences in histopathological response definitions. Excluding these studies a mixed effects logistic

regression indicated differences due to baseline uptake level (high vs. low repeatability) and treatment (CT vs. CRT). Summary receiver operating characteristic (SROC) analyses indicated that on studies and individuals with high baseline values (i.e. high repeatability) had better discrimination of responders than respectively CRT studies and individuals with low baseline values. For end-of-treatment (during treatment) FDG PET, a decrease of 65%-70% (40%-50%) provided maximal discrimination.

o **QIBA FDG-PET/CT Digital Reference Object Project** (Kinahan): Project complete. Summary: The QIBA FDG-PET/CT Digital Reference Object (DRO) provides a method for testing the SUV based calculations performed by PET/CT display software platforms. It is not intended as a ranking approach, but rather to verify nominal performance in a transparent and objective manner. The anecdotal evidence of failures (or otherwise) during evaluation of new software versions is a demonstration that the DRO has already succeeded in this manner. Even so, initial DRO tests have identified some variation in performance.

o **Analysis of SARC 11 Trial PET Data by PERCIST with Linkage to Clinical Outcomes** (Wahl): Project complete. Results: The median OS was 8.6 months (95% CI, 5.9 to 11.3 months) for the 115 patients. Using PERCIST, patients with progressive metabolic disease (PMD) showed shorter OS (median 4.7 months, 95% CI, 2.9 to 6.5 months) compared to patients without progression (median 10.0 months, 95% CI, 7.8 to 12.2 months). PMD on day 9 PET had a hazard ratio of 2.8 (95% CI, 1.5 to 5.5) for death. Changes in FDG uptake after 9 days of therapy was associated with survival at 1 year (AUC 0.71), progression at 3 months (AUC 0.63), and CT non-progression after 6 weeks of therapy (AUC 0.79). Conclusion: Treatment response by FDG PET as early as 9 days into IGF-1R antibody therapy in patients with ESFT can identify patients very likely to have poor outcome.

o **Evaluation of the Variability in Determination of Quantitative PET Parameters of Treatment Response across Performance Sites and Readers** (Wahl): Summary: IRB approval required at Johns Hopkins has been obtained, data has been anonymized, and sent to performance sites in the US and Europe. In the US, both CRO and academic sites have been identified. The QIBA technical committee sites and the QIN sites were given priority. The sites have initiated interpretation of these studies. We expect we will determine that the SUV max change measures are highly reproducible and expect an ICC of .90 to result. We expect repeatability among sites to be less robust for SUV peak and volumes of tumors. These data in which each site examines real human data and must select the "hottest" tumor for analysis with a variety of workstations, will allow us to determine what degree of variability is due to the reader, software and workstation, as opposed to being due to the intrinsic variability among test re test scans. These data will inform the QIBA community on what key gaps exist in our analytical approaches, thus helping drive our future studies in quantitative imaging.

o **PERCIST Validation** (Hoekstra): Project complete. Conclusions: Low to moderate heterogeneity in the PERCIST (single lesion) hazard ratios. In contrast there was minimal heterogeneity in the PERCIST (5 hottest lesions) Resp vs. SD hazard ratios, but very large variation between the PD vs. SD hazard ratios. With respect to prognostic value of the two PERCIST response rules, both response rules are equally and significantly prognostic for PFS (C-index), with the single lesion response rule being slightly more prognostic for OS. These results were not different in the sensitivity analyses. With respect to comparison of prognostic value the different response rules, all measures are equally prognostic, with the Langen response rule having slightly higher concordance index (of the order 0.02) and 5 hottest lesion PERCIST rule providing the least prognostic value for OS, however these differences are well within the variation of the indices (approximately 0.05). The merging of SD and PD groups in the PERCIST measures reduces their c-indices by around 0.03.

- o ***Evaluation of FDG-PET SUV Covariates, Metrics, and Response Criteria*** (Yap): Project complete. <u>Results</u>: 1. Compiled a research archive and database that supports FDG-PET imaging results, covariates, and clinical outcome data. 2. Developed software for importing previously defined tumor ROI and statistics files and generating additional metrics and statistics from the associated PET images (e.g., SUVpeak). This software has been applied to the results from a previous clinical trial of an experimental therapeutic (lapatinib) in patients with metastatic breast cancer. In addition, normal tissue ROIs have been defined for aorta and/or liver background regions to calculate additional tumor-to-background ratio (TBR) metrics. Statistical analysis has been performed to evaluate correlations between metrics, concordance between different response criteria (e.g. EORTC vs. PERCIST), and prediction of clinical outcome based on survival analysis.

- Profiling:
  - o The QIBA FDG-PET Profile for the therapeutic area of oncology is all but finalized for the public comment phase.

E. QUANTITATIVE MRI

Completed during the contract term:

- DCE-MRI, progress has been exceptional on all fronts:
  - o Experimental groundwork:
    - ***DCE-MRI Phantom Fabrication, Data Acquisition and Analysis, and Data Distribution*** (Jackson): The project is complete. Eight copies of the QIBA DCE-MRI phantom have been manufactured successfully by The Phantom Laboratory. Data from multiple MR scanner vendors have been acquired and analyzed.

    - ***Software Development for Analysis of QIBA DCE-MRI Phantom Data*** (Ashton): The work is complete. A number of software improvements and fixes have been implemented in the period.

    - ***Digital Reference Object for DCE-MRI Analysis Software Verification*** (Barboriak): Project complete. Accomplishments: We have succeeded in creating digital reference objects for both the T-1 mapping and dynamic portions of the DCE-MRI experiment. We are providing these as open-source objects, where both the imaging data and the methods used to create the data have been provided on our website. There has been considerable interest from software companies and academic sites in using these DROs to verify software packages. We have used the DROs created to begin the verification process of the open source software dcemris4 (www.dcemri.org). Because this software package is open-source, it is particularly suited to provide a reference data analysis method, and a basis for comparison to other available packages. We have begun the software evaluation process using the synthetic T1 mapping DRO. Over 15 software packages or software package variants have been evaluated using this DRO. As part of this process, we have undertaken a cooperative research project with ACRIN. In this project, a Princeton undergraduate Richard Price has contributed an automated method of summarizing and graphing the results in order to generate a report comparing the software packages as an aid for software evaluation.

    - ***Test-Retest Evaluation of Repeatability of DCE-MRI and DWI in Human Subjects*** (Rosen): <u>Summary</u>: Delays toward project completion were encountered, in part due to the needs for administrative re-organization within ACRIN in anticipation of the NCI NCTN group restructuring. However, participating sites have been actively engaged,

subject accrual is underway and, as planned, funding for ongoing activities for project completion is being undertaken by ECOG-ACRIN.

- o Profiling:
    - ▪ The MR Technical Committee has completed its first QIBA Profile on DCE-MRI and has changed its focus to diffusion-weighted MR imaging. Accordingly, the Technical Committee has renamed itself the Perfusion Diffusion and Flow MR Technical Committee (PDF-MR). This committee has a new co-chair, Marko Ivancevic, PhD. There is a one year timeline for developing the DWI profile, and a second version of the DCE-MRI profile has a two-year timeline. The DCE-MRI profile is not organ-specific; a new profile that is specific to the prostate will be developed for the ACRIN trial. The DCE-MRI profile has been implemented in some Roche clinical studies.

- The fMRI committee has developed provisional core details for a Profile and has defined tasks and approach to characterize reproducibility in the measurements.

    - o ***Quantitative Measures of fMRI Reproducibility for Pre-Surgical Planning-Development of Reproducibility Metrics*** (Voyvodic): Project complete. Results: AMPLE normalization did greatly increase quantitative reproducibility of fMRI language mapping, which compares standard t-maps and AMPLE maps for a single subject scanned 6 times. Quantitative measures of reproducibility showed that the brain location for the peak of activation in AMPLE-normalized maps was highly reproducible across scan sessions. We found that peak location of language areas varied by less than 10mm in almost all pair-wise comparisons, and by less than 5mm when both scans were acquired using the same scanner model and the same pulse sequence. Reproducibility of the spatial extent of activation was not correlated with acquisition procedures, but was positively correlated with the strength of the task activation signal itself. Reproducibility of hemispheric laterality index in AMPLE normalized maps was very good, with over 90% agreement between scans in frontal and temporoparietal language areas. The data sets used in this project could be made available for sharing via QIBA's Open Image Archive Initiative. New data sets should be acquired using a QIBA supported fMRI protocol, incorporating our recent findings, which could then provide publicly available support for the claims in our Profile.

    - o ***Quantitative Measures of fMRI Reproducibility for Pre-Surgical Planning*** (DeYoe): Project complete. <u>Discussion</u>: Overall, the results of this study indicate that visual cortex fMRI activation foci are, on the whole, quite reproducible with respect to the location of the weighted center-of-mass, the number of voxels within the focus and the spatial overlap of repeated observations. AMPLE normalization can improve repeatability relative to conventional T-valued data at 7 comparable statistical thresholds. However, anecdotal observations with other visual mapping data not presented here suggest that AMPLE normalization may at times suppress weak activation in portions of a broad pattern of activity (e.g. extrastriate cortex) that might be important for clinical interpretation in the context of presurgical planning. A manuscript for journal publication is in preparation that will describe additional details and results from this study that go beyond the items described above which most directly impact the claims being considered for inclusion in the QIBA fMRI profile.

    - o ***Validation of Breath Hold Task for Assessment of Cerebrovascular Responsiveness and Calibration of Language Activation Maps to Optimize Reproducibility*** (Pillai): Project complete. <u>Conclusion and future work</u>: The CVR calibration approach used in this sample dataset has demonstrated its potential to minimize false negatives for motor mapping in brain tumor patients with reduced CVR. A further development of the technique is needed to make it completely threshold independent and also to extend its application to mapping of eloquent cortex other than simply primary sensorimotor cortex in patients. Specifically, the technique

needs to be further evaluated and optimized for application to intrinsically less symmetric and more strongly lateralized functional networks in the human brain such as the more complex language network in patients with focal resectable brain lesions such as brain tumors. We have already shown that in normal volunteers this approach is useful for calibration of language activation maps, but in the setting of brain lesions, more work needs to be performed in order to establish this algorithm as a viable calibration/normalization approach for use in the clinical setting. Clinical validation is also needed via comparison with intraoperative electrophysiological mapping and correlation with patients' functional status.

F. ULTRASOUND/SWS

The ultrasound biomarker Technical Committee has been formed, and it has decided that sheer wave speed will be the first biomarker addressed.

For ultrasound, the challenges are well known. We need to develop an understanding of how to obtain comparable numbers across platforms. Three subcommittees have been formed:

- System dependencies-looking at how each system makes their measurements and the sources of variability
- Clinical applications- looking at making the measurement in humans and sources of variability
- Phantoms – designing phantoms to assess Shear Wave Speed measurement