QIN Annual Meeting
9609 Medical Center Drive, Rockville, MD; May 21-22, 2019

On May 21, 2019, Dr. Janet Eary, Associate Director of the NCI, CIP welcomed over 160 members of the Quantitative Imaging Network (QIN) and other stakeholders to their 9th Annual Meeting at the NCI Shady Grove Campus, in Rockville, MD. Dr. Robert Nordstrom, QIN Director, initiated the two-day program review by outlining the meeting agenda, the priority challenges for QIN tools translation into clinical applications and a roadmap for achieving these goals by increasing interactions with NCTN trial investigators, outreach to the RSNA Quantitative Biomarker Alliance (QIBA) and through efforts to commercialize various software and workflow products developed by QIN through the SBIR process. Currently there are over 67 QIN tools that have been produced by the Network membership, 13 of which were featured in demonstration displays during the meeting and evaluated by the Clinical Consulting Committee for readiness for clinical utility and trial deployment. To reach this stage of development, QIN tools have been categorized using a 5-step benchmarking process, 1) preliminary, 2) basic, 3) technically tested, 4) trial validated, and 5) demonstrating clinical utility.

Clinical Utility of QIN Tools

The keynote speaker, Dr. Daniel Sullivan, professor emeritus at Duke University Radiology, reflected on the history of development of quantitative imaging (QI) methods in clinical practice and in organized trials from the perspective of his career that has included both the leadership of the NCI, CIP as well as the founding scientific advisor for RSNA, QIBA. From the early beginnings where Geiger counters were crudely used to evaluate the ratio of thyroid iodine^{131} uptake over the neck versus a background somatic location in the thigh to the more sophisticated estimates of cardiac ejection fraction and fetal growth development monitoring from actual ultrasound images, the field of quantitative imaging has grown in both number of imaging modalities employed, and the complexity of quantitative methods utilized.

During the continued development of QI, challenges in areas of data sharing, consensus and standardization created a lack of focus in QI methods. For example, published surveys of how quantitative methods like SUV are employed in the calculation of PET response cite 13 different systems using 3 input variables. In the drug development space, the FDA has often expressed concern about the reliability and consistency of imaging methods in their use as trial endpoints, going so far as to retract the line extension approval of an anti-vascular breast cancer therapy where no independent blinded central review was conducted, and the findings could not be reproduced in additional follow-on studies. It is for these reasons, in late 2007, the RSNA decided to develop the Quantitative Imaging Biomarker Alliance (QIBA) program. Imaging experts were organized to "industrialize" imaging biomarkers by developing technical validation standards across the imaging modality sectors. This is being achieved through the generation of consensus Profiles that specify definitions, rigorous imaging acquisition methods, and training of actors/operators in order to ensure imaging data collected from multiple sites are comparable. The success of this approach has been proved by the generation of an FDA Guidance document

on QI in 2019 that cites the QIBA Profiles as sources to consider for approaches to standardizing QI endpoints for use in the drug approval process. Although the field is still confronted by challenges including resistance to employing QI and standardization quality control in both trial design and clinical practice including, and the perceptions that the QC needed for quantitation can delay trial recruitment and retard workflow as well as impact reimbursement, there are success stories like the value generated by cooperative groups like ECOG-ACRIN utilizing a well-accepted QI quality system in all of the trials it develops and conducts. Networks like QIN and QIBA need to continue to collaborate and proselytize for standardized QI protocols to advance imaging science and bring it to the bedside to improve patient care and accelerate drug development.

The QIN Perspective: Moving QI Tools and Methods into Clinical Trials and Clinical Practice

Dr. David Mankoff, University of Pennsylvania and outgoing Chair of the QIN Executive Committee spoke on the focus of the meeting being the deployment of well-advanced QIN tools into the clinic. He discussed how to increase outreach to the NCTN system, so the readiness and utility of the tools to aid in supporting clinical decisions making and trial endpoint analysis could be demonstrated.

To enhance clinical trial performance of QI tools, Dr. Mankoff mentioned robust automated segmentation assessment tools, better standardized target/non-target lesion metrics, residual disease identification methods, improved informatics and clinical annotation systems and imaging collection and curation approaches for all the above. These needs then must be matched with the appropriate clinical hypothesis a particular trial design is trying to answer. Several clinical areas such as how to enhance precision oncology by better characterizing and revealing tumor heterogeneity perhaps by "radiomics" approaches, by reducing treatment toxicity by employing more quantitative imaging monitoring methods, or by increasing the sensitivity and specificity of immune-oncology response prediction were presented. All these needs require adherence to the following steps 1) Analytic validation by defining the methodology and accurately measuring the output; 2) confirming the repeatability of the measurement by use of reference standards; and 3) using QA/QC routines to ensure calibration of the imaging instruments used. For examples, the recent publication of a consensus recommendation for use of QI methods for 18FDG-PET in oncology outline this stepwise process (Shankar et al). The report by Weber and ECOG-ACRIN investigators on the test/re-test methodology to demonstrate the confidence intervals for FDG-PET SUV data collected across multiple centers and scanners was key in defining the capabilities of this technology. The EA1142 trial where FES-PET was used to quantitate tumor ER expression employed a comparison to the gold standard IHC method as well as rigorous phantom calibration work. Early FMISO-PET radio-probe pilot studies imaging hypoxia in brain, head and neck and cervical cancer have led to the integration of this QI method into the ACRIN 6684 glioblastoma trial. Non-Hodgkin lymphoma studies now regularly integrate quantitative FDG-PET response prediction of CMR into trials such as EA2410. Finally, as previously described in his 2014 JNM paper, Dr. Mankoff reiterated that the distinction

between serving as an integral versus an integrated imaging biomarker is critical in determining the readiness of a QIN tool for use in a randomized clinical trial design. He finished by asking the volunteer Clinical Consulting Committee to consider these issues while observing the demonstrations of selected QIN tools and discussing their readiness and pathways for incorporation into select clinical trials.

Tool Demonstrations

The following Clinical Consulting Committee members then circulated around the various QIN tools demonstration screens and workstations to evaluate their readiness and convene later to provide a report to the QIN.

Janet Dancey, MD, FRCPS Queen's University
Freddy Escorcia MD, Ph.D. NIH Clinical Center
Daniel Sullivan, MD, Duke University
Michael Knopp MD, PhD Ohio State University
Larissa Korde, MD, Cancer Therapy Evaluation Program, NCI
Charles Kunos, MD PhD Cancer Therapy Evaluation Program, NCI
Anthony Shields, MD, PhD Wayne State University
Edward Jackson, PhD University of Wisconsin
Susan Chang, MD University of California at San Francisco

Other meeting participants adjourned to attend the following QIN Working Group (WG) breakout sessions including PET/CT; MRI; Bioinformatics and Data Sharing (BIDS); or Clinical Trial Design and Development or took the opportunity to view the Poster Session. Reports from the WG discussions were also presented later in the meeting.

Review of the Co-Clinical Trial Meeting from May 20

Dr. Huiming Zhang, NCI-CIP Program Director introduced the topic of preclinical QI and the co-clinical trial cancer resource program. The goals of this initiative include demonstrations of how genetically-engineered mouse models (GEMM) and patient-derived xenograft (PDX) animal systems can help develop QI methods whose information can be made available through a web-based resource hub. She introduced the following speakers to describe their work in the program:

Dr. Kooresh Shoghi, Washington University, related efforts to harmonize clinical and preclinical scanners by the use of co-clinical phantom calibration, test/re-test data gathering and comparisons to tissue pathology gold standards in PDX models of TNBC neoadjuvant therapies. Employing both FDG-PET and MR imaging, this group has explored the heterogeneity of PDX samples both in terms of phenotype and growth profiles, by optimizing noise/resolution voxel size in preclinical radiomics, producing an atlas of H&E pathology vs PET features and their relationship to tumor biology, and examined Apparent Diffusion Coefficient (ADC) MRI signal in comparison to PET-detected tumor metabolism.

Dr. Cristian Badea, showed the Duke University preclinical program's use of a GEMM flox/flox inducible sarcoma model that at 8-12 weeks develops a tumor profile that can be treated comparably to the regimen in an ongoing Phase II PDL/adjuvant radiation therapy trial. Both clinical and preclinical designs can use CT and MRI to monitor the development of metastases after primary tumor resection and adjuvant treatment. The group has performed investigations into phantom calibration in determining the need for surface coil correction of both T1and T2 imaging with and without contrast, the influence of limb positioning on the accuracy of semi-automated segmentation of tumor volumetry, the value of pocket phantoms in assessing the impact of respiratory gating on lung tumor nodule detection, and have participated in data sharing testing of preclinical segmentation protocols with other programs in the network.

Dr. Charles Manning described the activities of the Vanderbilt University "PREDICT" program. This U24 project is focused on the development of new PET tracers in collaboration with the VU GI SPORE. Industry developers are optimizing probes that can be used to detect both indeterminate lung nodules and visualize hepatocellular, colorectal and pancreatic carcinoma in human immune system PDX and GEMM models. In the former, a Ga68 folate tracer EC2115 provided by Endocyte Inc. has been used to stage COPD and identify lung tumors. In-house development of both F^{18} and C^{11} acetate and glutamate investigational tracers have been used to assess preclinical tumor metabolism and mechanism of action in response to therapeutic treatments as well as evaluate their readiness for incorporation into monitoring target engagement of EGFR inhibitors (cetuximab) or ABC transporter inhibitors (V-9302) under actual human study conditions.

Dr. Rong Zhou, provided details on the University of Pennsylvania Quantitative MRI Resource Program for Pancreatic Cancer. Pancreatic cancer is unique in the production of a dense stromal capsule consisting of both collagen and hyaluronate surrounding the individual epithelial lesions. This results in a high interstitial pressure resisting the vascular access to chemotherapeutic drugs. The Penn program is studying Dynamic Contrast-Enhanced (DCE) MR imaging as a preclinical QI marker of the impact of a PEGylated hyaluronidase (PH20) on tumor response using both a GEMM Kras/p53, PDX, and syngeneic murine pancreatic cancer orthotopic implant models. Validation of the 24hr. pharmacokinetics of injected PH20 is modeled using both AIF blood flow and K-trans methods and validated against an IHC hyaluronate staining standard. Other research activities include optimization of spatial-temporal resolution control using respiratory gating and the study of various radiomics applications using these systems.

Young Scientist Award

Dr. Nordstrom introduced the recipient of the Larry Clarke Young Investigator Award, Dr. Saumya Gurbani from Emory University who presented an overview of the Brain Imaging Collaboration Suite (BrICS) developed within his QIN program. The focus of his research has been the development of a machine-learning algorithm to improve the interpretation and clinical translatability of MR spectroscopy (MRS) in aid of glioblastoma (GBM) treatment. GBM tumors are some of the most aggressive and refractory to treatment of all cancers with 10,000 cases in the US annually demonstrating a meager 15 month median overall survival (OS) time. The standard of care management for this tumor surgical resection, followed by radiotherapy and

then chemotherapy has not improved this OS rate in several decades with most patients experiencing a recurrence in six months. The standard imaging procedure for charting the extent of GBM lesions and planning XRT therapy has been T1-weighted MRI using contrast that delineates the margins of the tumor given its disruption of the blood-brain barrier permitting leakage or flair of the contrast to distinguish the tumor boarder and associated edema and inflammation from normal brain parenchyma. Typically, image-guided radiation doses are directed at higher levels to regions of high contrast with lower radiation doses to low contrast segments of a lesion. However, questions remain regarding the accuracy of this imaging paradigm in delineating the margins of the tumor and how to improve delivery of therapy and patient survival. MRS technology has been studied over the year as a way to quantitate the differences in the chemical signature of normal tissue from that of a tumor. It has been shown that the ratio of the MR spectrum of choline and acetyl aspartate (CHO/NAA) can distinguish between healthy neuronal and white matter membranes and cytoplasm and those of GBM. The Emory group took on the challenge of improving on the MRS commercial software tools on the market including MIDA LC Model, and better integrating them into the clinical patient treatment management workflow. The improved QIN tool evaluation took place in the context of a three center MRS-guided XRT trial conducted at JHU, Emory, and Univ Miami that has now enrolled 40 subjects (29 treated to date). The novel software enables extension of the tumor margins treated with high or low dose XRT to the regions displayed by an MRS CHO/NAA ratio pathologic threshold overlaid onto standard MR images in the experimental therapy arm, with comparisons of outcomes achieved by radiation planning derived from standard MRI images. Twelve–month follow-up for survival and assessment for quality of life impact on neurocognition resulting from the wider field of radiotherapy and area of resection by the experimental procedure are endpoints in this proof-of-concept study.

Commercialization Pathways

Dr. Greg Evans, and Dr. Deepa Narayanan from the NCI Small Business Innovation Research (SBIR) Development Center reviewed how the SBIR/STTR programs might aid QIN tool developers in obtaining funding and commercialization expertise to help bring their imaging product concept to market. They pointed out that the overall scope of the SBIR/STTR budget at 3.65% of the overall NCI budget is available for eligible applicants in small businesses of <500 employees at least >50% located in the US. Funded investigators for SBIR grants need to be at least 60% supported by a company position, whereas STTR grants permit academic innovators to be still 60% supported by their university salary. The prototypical Phase I budget is in the $400K range and is awarded for 6mos to 1 year, while Phase II support, which requires a commercialization plan, can run for up to 2 years with a $2M budget maximum. A successful application should emphasize the competitive advantage of the product candidate over the existing market. Technical assistance support for external consultants such as regulatory affairs or manufacturing specialists can be incorporated into the budgets. At present the NCI SBIR/STTR portfolio contains over 450 companies of which ~15% are developing imaging technologies. The omnibus solicitation period for grant applications takes place thrice yearly in September, January and April. Other funding mechanisms such as Phase IIb bridge grants (that

require majority outside private investor support) to help enable the last hurdle of clinical or device approval are explained in greater detail in the NCI SBIR/STTR website. The SBIR staff then took questions from the meeting audience, followed by the QIN Tools Commercialization Panel discussion.

- There are already Academic/Industrial Partnership Programs supported by NCI-CIP. How are these different? A: Larger companies can participate in these programs not just small businesses.
- What is the ratio of SBIR to STTR funding? A: No typical year, sometimes one program has more awards, sometimes the other.
- How are topics nominated for Omnibus solicitations? A: NCI staff generates them for their institute. Advocate them to your Project Officer.

The QIN Tools Commercialization Panel consisted of the following small business representatives:

- Eman Nemate- 9-point Medical (product focus: optical coherence tomography to detect Barrett's esophagus dysplasia)
- Trinity…- Novometrics (product focus: clinical trials workflow software-Mass General technology)
- Antonine Gumbari- Kitware (product focus: open-source imaging software; eg., 3D-slicer)
- Jay Odopa- Univ Penn (product focus: Quantitative Radiation Solutions; autocontouring/segmentation software for XRT in H&N and thorax

The Panel was asked to reflect and comment on the following questions. Examples of some of the Panel's responses are included.

- Why commercialize QIN tools? A: Single institutional research needs validation in a larger environment. Often customer's discovery issues with products and user feedback is crucial to improve initial designs.
- What QIN product features make for stronger commercialization candidates? A: Upfront standardization awareness and considerations make for strongest designs; Emphasis on quality management system most important
- Based on your experience what are the greatest hurdles to commercialization? A: Institutional IP position has to be clarified or there will be problems later; the need to move to market quickly or encounter problems with the currency of technology
- Can an academic software developer be directly involved in company commercialization? A: Yes/No, see some of the SBIR/STTR requirements; depends on institutional rules
- Can small businesses play a role in deploying QIN tools in clinical trials? A: Yes get suggestions for customization of open source platforms from other small business users; get evidence of small business adoption to advocate/market for trial incorporation

- At what point in the commercialization pathway should a developer have a conversation with the FDA? A: Pre-submission conferences available and very valuable; 3-month lead time for information meetings; 12-month lead time for 510K/PMA submissions

QIBA – QIN Interactions

Dr. Ed Jackson, University of Wisconsin, current Chairman: RSNA, Quantitative Imaging Biomarker Alliance (QIBA) covered the rationale for the founding of QIBA, their work activities over the years and opportunities for future interactions between QIN and QIBA to further the adoption of quantitative imaging in various medical fields in the future. The key objectives of the various QIBA working groups in all of the major imaging modalities are to identify sources of bias and variance in imaging acquisition, patient preparation, and analytic techniques such as use of reconstruction algorithms and reduce their impact on the accuracy and precision of QI. The primary tool used by QIBA WG is to author a consensus document for each imaging application called a Profile, that identifies all critical actors in producing the image and recommends best practices and conformance approaches to harmonize standard operations in the generating the images and stating CI% whereby QI metrics can be reliably and reproducibly reported in both cross-sectional and longitudinal settings. Financial and scientific manpower support to collect the evidence to help confirm these Profile claims for imaging methods such as PET, CT, MRI and US, have come from RSNA, NIBIB, NIST, FDA, NCI, the imaging CRO industry and many academic centers. Projects have looked at reader concordance, phantom development, and helped codify metrology lexicons and statistical considerations for best QI analysis approaches. Ongoing collaborations are sought to clinically confirm many of the Profile claims that are undergoing technical validation. A demonstration of the impact that QIBA has had on contribution in the QI field has been the inclusion of a description of the QIBA Profile process, in the recent FDA Guidance on Imaging Endpoints in Clinical trials, as a valuable resource for sponsors considering how to incorporate QI methods into their drug development and approval programs.

A panel discussion was conducted among a number of participants that hold membership in both QIN and QIBA on how more interactions between the groups could be fostered. Comments included:

- The Emory brain imaging, and the I-SPY2 breast cancer imaging studies could provide ideal settings in which to demonstrate QIBA CT/MRI volume Profile validity
- The ongoing EA1183 breast cancer bone metastasis study used the QIBA FDG-PET Profile as the basis for protocol development and is using the AutoPERCIST QIN tool to assess response.
- Physical phantom calibration work advocated by QIBA Profiles is being used in many QI studies. Data from QIN PET segmentation tools can also be standardized using a QIBA developed PET digital reference object (DRO)
- Better monitoring of compliance with standards promulgated by guidance documents like Profiles should come from professional societies. Examples from the clinical pathology

world include accreditation review for use of structured pathology reporting could be emulated in the imaging field

- The basic principles of how to better advocate embedding QIN tools in more clinical trials include: 1) Reach out to the PIs when protocols are being designed; 2) Don't make imaging studies and CRFs so complicated they impinge on the study workflow; 3) Consider the need for core/central lab adoption of the QIN tool vs easy enough to be adopted by site readers
- Can IROC serve as a clearing house to advocate particular QIN tools in upcoming NCTN trials? Will early hand-off of tool for commercialization promote assessment in more trials? Is it valuable to incorporate tool elements into DICOM standard to enhance vendor adoption and interoperability?

Administrative Announcements - Robert J. Nordstrom, Ph.D.

Dr. Nordstrom reminded the attendees that phantoms used in QIN were displayed in the poster presentation room, and encourage attendees to take a look what these objects look like. He also reminded the attendees that QIN has been successful in creating issues of journal completely dedicated to QIN research with the most recent being the Tomography journal published in March 2019. QIN was encouraged by the journal to consider creating another issue, dedicated to quantitative imaging. The QIN Executive Committee meeting in the summer will discuss the feasibility of such an issue. Session 7 was designed to be a Q/A to hear feedback from the Clinical Consulting Committee from their review of the QIN tools demonstrated yesterday and conversation with the investigators. Members of the Clinical Consulting Committee are clinicians QIN invited to provide clinical perspectives on the tools. We invited Dr. Gary Kelloff to lead the discussion to provide unbiased thoughts on the paths of the QIN tool development. The Session started with a panel discussion followed with Q/A.

QIN – NCTN Tool Developer Discussion

Follow-up to the Clinical Consulting Committee session
Discussion and feedback on Clinical Consulting Committee presentation; Strategy planning on next-steps for QIN tool translation.

Chair:	Gary Kelloff, MD, NCI	(GK)
Panelists:	Alliance: Larry Schwartz, MD, Columbia University	(LS)
	ECOG-ACRIN: David Mankoff, MD, PhD, U Penn	(DM)
	NRG: John Buatti, MD, University of Iowa	(JB)
	IROC – Mike Knopp, MD, PhD, Ohio State University	(MK)

Dr. Gary Kelloff, NCI, led the panel discussion. He started with asking specific questions to the panelists, and then opened the session to group discussion.

Assume some of the tools are "shovel ready" and some are not. We would like to look at QIN/QIBA/NCTN to get a landscape analysis to understand what opportunities exist in the ongoing trials to test QIN tools. (GK)

The imaging community should employ various approaches to promote testing and validation of their quantitative imaging (QI) tools.

- We should identify pre-existing datasets that are archived, such as those in TCIA, and those that have been made available through ECOG-ACRIN and QIN U01s. (DM)
- We should continue to expand engagement with clinical trial investigators by attending scientific meetings and to understand what trials are being conducted and to be conducted in different trial groups. The culture of clinical trials which centers on demonstration of treatment effects of therapeutic interventions, and the value of quantitative imaging in helping clinical trials is not apparent to clinical investigators. The emphasis should be put on getting clinical oncologists engaged with imaging committees in the cooperative groups. (DM, JB)
- The QIN "shovel ready" (SR) tools can be tested prospectively in clinical trials in a hypothesis driven fashion in order to have real clinical impact. These tools can be tested as integrated markers as secondary or exploratory endpoints. ECOG-ACRIN is in the process of testing QIN tools as secondary or exploratory endpoints in several trials, including a bone metastasis-dominant breast cancer study with FDG-PET/CT using auto-PERCIST. (DM)
- None of the current QIN tools are mature enough as an integral marker in prospective trials as predictive or response markers. (DM)
- To help match the QIN tools with clinical trials, IROC created a questionnaire for clinical investigators to fill out at the time of a clinical trial concept formation to understand the radiotherapy (RT) and imaging modalities that will be used in the trials. But the questionnaire/process should be designed in a way not being viewed as an extra burden to the investigators to encourage participation. (IROC, DM).
- For the shovel ready tools, we can propose a couple of pilot projects to structurally walk through the path of clinical validation to identify obstacles, including the volume data and required informatics infrastructure. (CIP, Yantian Zhang)
- It is important for the QI tool developers to apply their tools to the most recently collected real world data so that they do not underestimated the variability of the real world data. (MK)
- It is encouraging to see in neuro-oncology that there are rich data sets from multiple trials that can serve the purpose of assessing the feasibility of QI tools. (MK)
- Combining data from many different trials will help understand if additional specifications are necessary to test the QI tools. (MK)

Barriers (financial and logistic) of adding QI biomarkers to prospective trials

- Decisions on adding an exploratory or secondary endpoint in many cases are financially based because of the associated actual or perceived burden to patient enrolment. However, there may be a niche that allows one to help answer the prospective question with an imaging biomarker without incurring unacceptable costs. (GK)
- In cases where standard of care (SOC) images are adequate for QI markers, the cost may not be prohibitive because the images are paid for by SOC (*e.g.*, clinically indicated CT,

MRI, FDG-PET). No additional cost is required for analyzing the data as most of the QIN groups are funded by existing mechanisms. (DM)

You mentioned that progress will be made on a case-by-case basis. As a committee chair for both Alliance and SWOG, what do you consider to be the lowest hanging fruit in terms of either target organ or stage of disease? What do you think are the shovel ready tools and where are they tested now? (GK)

- The low hanging fruit for testing and validating QI tools is the existence of unmet needs, that is, in the disease states where conventional tools for response assessment don't work. Imagers should work with oncology colleagues to identify these areas. The imaging communities at Alliance, SWOG, ECOG-ACRIN, NRG, COG etc. can be helpful in bringing those opportunities to the attention of QIN and QIBA, and vice versa, we alert them to the value of some of the QIN tools. (LS, DM)
- In many settings, SOC imaging or the analysis of SOC imaging may be suboptimal. We have to better define what the unmet need is in clinical trials or clinical practice. Is that the analytic assays being better, or is that a more nuanced response assessment metric that may involve more than a binary, positive or negative assignment? (LS)
- The most productive scenario is when a disease committee chair in cooperative groups comes to imagers and says that I have this trial that needs novel imaging tools. (LS)

The SOC imaging is being covered in a lot of trials, but is there a need for resources to cover aspects that aren't SOC to obtain data that are needed for QI tool assessment? In that case, IROC could be an honest broker to involve the help from SBIR to further develop the QI tools. (GK)

- Resources and support are most needed where non-SOC time points for imaging assessment need to be collected in order to test and validate QI tools. (MK, LS)
- From IROC's experience in managing NCTN trials, the more we understand the important parameters for the imaging tools, the more confident we are that we can assess the feasibility and the variability of their utilization.
- Standardizing the nomenclature is very important, for example, for AI algorithms and feature descriptors.
- QIN teams are encouraged to spend effort in collecting and saving imaging and clinical annotations. When the images with associated annotations become part of sharable archives they are more valuable. This has been done on a smaller scale for some QIN challenges, and could be expanded to a larger scale so that we can have a large annotated data sets available that developers can more readily use to assess their QI tools.

Importance of high quality imaging acquisition and processing

- The images from SOC procedures should be of high quality in order for them to be useful for QI tool assessment. We should also develop easily useable and robust tools that provide calibration and qualification for standardized acquisition, in particularly tools that are applicable to study centers across the board, so that not only data from academic centers but also data from non-academic centers can be of high quality and interpreted consistently. This consistency is especially important for MRI and PET because these imaging modalities have high variability. (DM)
- Imaging quality is extremely important for QI tools to be successful. When we are pushing certain quality standards, there is logically the pushback "show me that it is impacting

things like interpretation". QIBA is extremely valuable in setting and promoting quality standards. But the impact of measurement variability is something that we as a community have not publicized enough. As we start to publicize this, we have all the reasons to start ratcheting-up expectations on the impact of imaging quality. (MK)

- The better the confidence in the imaging quality, the better the confidence on the call of treatment effect. Then we can gather more support from pharma for imaging-based biomarkers. (GK)
- Imaging segmentation is the fundamental image processing operation that is required for any kind of quantification. Segmentation as a function of the imaging quality has not been studied. It is important to establish the performance of segmentation algorithms as a function of imaging quality.

From the NCORP perspective, what is the lowest hanging fruit of QI tools that radiation oncologists can use? (GK)

- First and foremost, we need standardized algorithmic approaches to define the target for radiation therapy (JB) to reduce intra- and inter-operator variability so that tumor control can be accurately accessed. Target definition is currently done manually with high levels of inconsistency and is unsustainable. We need to conduct trials to test and validate QI tools for tumor target definition. (JB)
- We have a real opportunity for QI tools to show value therapeutically in the field of theranostics. Segmentation and dosimetry tools that enable personalized dosimetry for dose-adjustment to limit normal tissue exposure and maximize tumor radiation dose are needed to optimize radiation therapy. (JB)

Standardize terms to describe the levels of the maturity of QI tools.

- The terms "mature" and "shovel ready" are too vague and mean different things to different people. The terminology of "integrated" and "integral" is familiar to NCTN/CTEP. QIN could consider adopting them and use them more consistently.
- The CIP staff have produced a spreadsheet of the QIN tools. The next steps would be to have an internal peer-review vetting process to determine which of these are ready for which level of testing in NCTN/CTEP. We need to make it easier for everyone to understand what is ready for integrated or integral testing. The additional advantage of going through this process could be that associated funding would go to support the development of these programs. (DM)
- The determination of integrated or integral classification falls on the imaging science community. Because without that knowledge or those standards, it's going to be difficult for anybody to understand the utility or how the QI marker would be classified. (GK)
- This may be an area where QIN could work together with NCTN and NCI staff to produce a White Paper which could have high impact. This will help both imagers and oncologists to understand the context in which these imaging biomarkers can be tested or utilized. (DM)
- The level of maturity of a tool (shovel ready) is relative to the clinical needs. So if there's a strong clinical need, we may have some tools that may not be as ready as others, but they may be even more important to include. Or if we have a shovel ready tool that doesn't have a strong clinical need, there may be more resistance. The two (tool and clinical need)

have to be matched up on an individual case-by-case basis, disease-by-disease, and that's what's going to get the tools out to implementation outside QIN. (LS)

What is the collective opinion about the opportunities for QI tools in different stages of cancers? (GK)

- With advanced computational capability, we can now look beyond the classical approach that we have been using for decades, which is we have an organ disease and we want to do specific imaging for this organ. We can look at the disease more comprehensively by combining imaging and other information (*e.g.*, liquid biopsy and clinical data) in more intelligent way to assess patient risk, and use risk-adaptive targeted approaches for screening, which minimizes the burden on low-risk patients without compromising a successful detection. (MK)

Besides NCTN, NCI has a program for earlier development, ETCTN. It tends to support trials which are smaller and sometimes not randomized. Is ETCTN better suited for testing QIN tools in some cases? (GK)

- ETCTN is interesting because a trial sometimes has a couple of sites, relatively small, but has a relative openness. If it doesn't burden the trial, we may try some exploratory processes to gain insight into a range of questions from the mechanism of action to expected toxicities. In contrast, I think the shovel ready tools will probably need to go to NCTN. The tools that may not be shovel ready could be implemented in a couple of sites in ETCTN to determine feasibility. (DM)
- From my perspective there wouldn't be a problem including an imaging question in smaller multicenter trials as a development strategy. It follows biomarker development and or even drug development paradigms. It is really a question of identifying who among the imagers would be part of the willing coalition to work on those aspects. (Janet Eary)
- For tool developers, ETCTN is especially interesting because the trial itself has a shorter duration. That is really important because you want to have your tool rapidly associated with a clinical publication. Some of the larger NCTN trials will collect data for 3–4 years, whereas ETCTN trials may have a significantly shorter duration for data collection. (MK)
- We can also leverage NIH funding from Specialized Programs of Research Excellence (SPOREs), which require a translation endpoint. Many SPOREs have early phase trials that may not be ETCTN studies. Exploratory endpoints can be placed to add correlatives imaging assessments, especially using the QIN tools. There is a recent JCO publication from a SPOREs study (TBCRC026) in HER-2 positive breast cancer that QIN investigator Dr. Rich Wahl is involved with. (comment from audience and DM)

Imaging acquisition parameters that are not part of standard DICOM headers: For the parameters that are important for QI but are not in the standard DICOM headers, such as MRI diffusion sequences, how could we work with scanner vendors to have the parameters recorded for future extraction and use in developing QI tools? (Question from audience)

- This is a constant dialogue with the scanner vendors. We are encouraging them to put as much information into the private tags that we know how to decipher. The next important question is how we standardize this because each scanner vendor has a different approach. The first step is that we are recording and defining aspects that are vendor-specific. Secondly, it is engaging with the vendors to understand how we can capture the

information. They are in general receptive to putting non-standard parameters in private tags. This is part of open engagement and can be achieved through different organizations like QIBA and others which have vendor participation in their activities. (MK)

- TCIA has a unique advantage at the moment and is happy to relay such requests for inclusion of such information in the private tags to the vendors. We maintain a knowledge base of where to find scientifically valid information in private tags, and one of the QIN chairs is in the DICOM standards committee. So we have a unique ability to address all of the vendors at once. (TCIA)
- QIBA has a history of engaging scanner manufacturers. Inclusion of the information in the private tag can be a good spot for the engagement as well. (Paul Kinahan).

QI tools in optimizing theranostics

- It's intuitively obvious that dosimetry will lead to more accurate delivery and hopefully better outcomes, but nobody has shown that in theranostics. The prevailing approach in theranostics is one size fits all, with the same dose, the same dosing schedule and duration of treatment. If dosimetry is more accurate, we can direct higher dose to the tumor, and lower dose to normal tissues. This will lead to better outcome. QI will be critical to test this hypothesis in a clinical trial.
- For therapeutic radiopharmaceuticals, we are under-exploiting the theranostic possibilities with the one size-fits-all dosing constraint that is approved by the FDA. Dosimetry could be an integral marker and may help optimize dosing for each individual patient, but is not on the drug label. (comment from audience). Such studies may need to be done under an IND. (DM)

Imaging quality and evidence on QI affecting outcomes

- When we promote better imaging quality and quantification to payers and other organizations, the pushback is: "How does it affect outcomes? Show us the data that it affects outcomes". So in the real world the issue is that performance of the tool alone is not enough. We do need to have the data to show that QI tools affect outcomes that make a difference for payers. (D Sullivan, LS)
- Outcomes-based studies of QI tools can be done retrospectively on archived data on TCIA and IROC and it would be very helpful and if QIN can generate some funds to do that. (D Sullivan)

Data Science, Pipelines & Radiomics

Radiomics Pipeline Demonstration & Discussion

Dr. Sandy Napel, Stanford University, described the Stanford Quantitative Imaging Feature Pipeline (QIFP). The pipeline is a Linux server-based workflow management system. It contains several Docker containerized modules performing various tasks, including intake of imaging data, imaging segmentation, radiomic feature extraction, construction of predictive models using radiomic features and clinical data, and output of a predictive QI biomarker based on the best combinations of features with clinical data.

He noted that the pipeline contains pre-configured workflows. Currently radiomic feature engines run in the pipeline include pyRadiomics, SIFT, and Delta-features. It accepts Docker containers from

users, including containerized machine-learning modules. Researchers can also configure their own workflows using Docker tools from other sources.

He then demonstrated how to use this workflow for discovery of a predictive biomarker in the QIN Bioinformatics/IT and Data Sharing Working Group (BIDS) project. They assembled a pipeline using containerized tools from several institutions, and ran that pipeline on public data (SPIE lung CT cohort with seed points for segmentation). The imaging data were imported in the QIFP pipeline, segmented by a segmentation tool in a Docker container. The radiomics features were then extracted by one of the three radiomic feature engines, and were used to build a predictive model based on 60 training cases to discover the best combinations of features to construct a biomarker. This biomarker was validated on the validation dataset of 10 cases.

Ashish Sharma, PhD, Emory University, described the use Google cloud computing environment to run Cloudy Pipelines, which is created at Emory University to scale up the QIFP pipeline for validation of machine-learning (ML) algorithms to advance the pipeline to the stage of production deployment (shovel ready). They used the Stanford QIFP pipeline structure and containers, where individual algorithm tools that performer various tasks are containerized. They describe the workflow in Workflow Description Language (WDL). When a WDL file is provided to the cloud environment, it triggers a series of events: the application programming interface (API) pulls in data, launches parallel pipelines, gathers results and notifies users. The users can download the data or model for this computation. The cost of running the training pipeline is about $1, and that of the testing/validation pipeline is about $0.25. The time required is comparable to that run on QIFP. This scaled-up cloud computing capacity has been tested on pathology images. In the past six months, they have used Cloudy Pipelines to process 3000 pathology images. They are managing an ongoing PhysioNet/Computing in Cardiology Challenges (CinC) challenge where participants submit algorithms as Docker files, and the Cloudy Pipelines run these algorithms on private real-world test dataset within the environment. They have received 200 algorithm submissions from 50 participants since mid-April. Bringing algorithms to data makes algorithm development much more efficient, and allows algorithm developers to test/validate their algorithms on sequestered or private data which would otherwise be unavailable to them.

<u>Panel Discussion Topics: Tool Development and Dissemination</u>
Panelists: Sandy Napel, Tom Casavant, Dan Rubin, Ashish Sharma

- First step would be to have tools containerized. Can QIN provide instructions on website on how to containerize tools?
    - It is relatively easy to get a tool in a Docker container.
- What is the advantage of putting a tool in a Docker container
    - It can be plugged in the existing pipeline and see what effect it has on the performance of the overall pipeline. The developer has control over who is allowed to use it.
- What are the additional incentives to containerizing tools and for others to test?
    - Tools can be tested on a wide range of data to make the tools more robust, and can be deployed to other cancer registries/datasets to make them more broadly applicable to real world situations.

- Standardization on data formats for the inputs and the outputs at each of the stages/modules of the pipeline.
    - For images, the most widely use formats are DICOM and NIFTI for data inputs. DICOM SR TID 1500 has been adopted by a number of tools as the output format. DICOM format has the advantage to interface with clinical data and clinical practice.
    - A critical next step is that the community needs to converge on data standards for both imaging and clinical data to achieve interoperability.
    - Opinion against DICOM: DICOM is the most cumbersome format for algorithm developers, and it is harder to work with. Neuron imaging community has settled on Brain Imaging Data Structure (BIDS) standard file format. It is more flexible. QIN is encouraged to look at it.
- There is also a need for standardization of ontology of radiomic features.
- There are many parameters in radiomic features and many classification methods. In practical implementation, it is a balance between containerization (fixed module) and leaving room for flexibility to optimize the algorithm.
- The QIN pipeline is generic and has interoperability for modules that is accessible to the community. Additional activities can be under this umbrella.
- What are the values of pipelines?
    - Clinical practice and decision making is all based on pipeline/workflow.
    - For tool developers, they can mix and match components, and compare the outputs.
    - They allow one to research higher impact products (modules) in the context of a decision-making workflow, from start to end, in the system already exist, so they can be fit in clinical settings readily.
- Is there quality assurance in place on high fidelity of the pipeline?
    - Quality assurance was not part of the pipeline design consideration.
    - The pipeline is not intended for solving validation problems, rather it allows interchange of the parts to performance comparison of modules to go through iteration process faster.

<u>Standardization in Quantitative Imaging: A Multi-center Comparison</u>

Lubomir Hadjiyski, Ph.D. Univ. of Michigan, presented the work from a collaborative project within the QIN PET/CT subgroup, led by Dr. Mike McNitt-Gray of UCLA and participated by nine sites/members of the subgroup. The project was designed in response to the need for quantitative evaluation and standardization of radiomic features. The long term goals are to develop imaging biomarker quantitative standardization tool, to estimate the difference of a newly developed feature from the existing set of features in an ontology, to determine if the proposed feature is a new feature type, and to classify the feature within already existing ontology. This tool will be made publicly accessible for the community so others can use it to check their radiomic features.

They used the International Biomarker Standardization Initiative (IBSI) reference manual as the source of feature definitions to investigate the agreement among radiomic features when computed by several groups utilizing different software packages with standardized feature definitions and common image datasets designed to identify possible differences.

Nine common quantitative imaging features were selected for comparison including nine features that describe morphology, intensity, shape and texture. These features are tumor volume, surface area, 2D diameter, 3D diameter, sphericity, intensity mean, intensity standard deviation, intensity kurtosis and gray level co-occurrence matrix (GLCM) entropy. The common image data sets were: (a) two sets of 3D Digital Reference Objects (DROs) developed specifically for this effort (200mm and 50mm diameter objects): a uniform sphere, a sphere with intensity variations, and a complex shape object with uniform intensity; and (b) 10 patient image scans from the Lung Image Database Consortium image collection (LIDC) dataset using a specific lesion in each scan. For DROs, six of the nine features demonstrated excellent agreement among different tools with CV < 1%. Larger variations (CV≥ 13%) were observed for the remaining three features, namely surface area, sphericity, and GLCM entropy. A similar trend was observed when the tools were applied to patient data. The working group will present the results at the upcoming AAPM conference, and will ultimately generate a manuscript for peer-reviewed publication. They have completed several debugging activities, and are working to understand the reasons behind the large variability in surface area and GLCM entropy.

Questions and comments:

- Are you also aiming to look at pre-processing? Sometimes pre-processing can cause variability as large as feature implementation itself.
    - We have not prescribed pre-processing specifications, but we understand it affects the downstream events to certain degrees. However, we documented the pre-processing differences.
- There are many different ways of interpolation. How did you estimate the surface area, based on voxels or triangulation?
    - We did not defined a specific approach; each site chose their own methods. The purpose is not to limit variation but to evaluate the differences from various methods.
    - This open approach is beneficial for debugging.

MRI Radiomics CCP

Dr. Russek, NIST, presented a collaborative project between QIN MRI Working Group and NIST on designing and constructing a biomimetric phantom for radiomic and AI image analysis. The project started a few months ago. The role of NIST in this project is to construct a biomimetic phantom that matches several tissue parameters and has realistic morphology and inhomogeneity to help validate radiomics in AI imaging analysis.

He reviewed NIST's long history of developing phantoms as reference objects. For a typical phantom, it is desirable to have stable materials, simple properties and geometries. In contrast, for a phantom to be used in radiomics, the phantom materials will need to have complex properties that mimic tissues, and also complex geometry. The proposed brain phantom contains a synthetic tumor that has all the complexities in both material parameters and geometry that you find in the real tumor and will be embedded in realistic brain tissue. The reason for the required complexities is that radiomic features are very sensitive to image protocols, and there is a need to make radiomic analysis independent of scanner by separating variability introduced by the scanner from that of the object/tumor. Additionally, image reconstruction algorithms may change radiomic features (not one-to-one correspondence between sensor space data and image data).

Therefore we need an object as a reference to understand what information is being kept in the reconstructed images, and what has been lost. MRI is much more complex than CT in that the image is a function of many tissue parameters, and there is a lack of consensus on the robustness of radiomic analysis. All the above point to the need of a precise MRI reference object.

To construct a phantom to match tissue spin relaxation times (T1, T2), proton density (PD), apparent diffusion coefficient (ADC), density, conductivity, and dielectric constant, they chose hydrogels as the material, which can be UV cured and make highly stable complex and heterogeneous structures. They hope to mimic all properties of soft tissue with various types of doped hydrogels by modifying both the polymer properties and internal doping to match MR resonance properties as well as electromagnetic properties. They are co-processing tissue and tissue-mimic measurements using identical measurement protocols to measure tissue and tissue mimics simultaneously.

The proposed plan is to develop a numerical phantom with synthetic tumors using realistic tissue growth algorithms, convert numerical phantom into 3D-printable output (stl, gcode), develop 3D printable tissue mimics and multicomponent bio printers, and finally disseminate phantoms to evaluate radiomic analyses. This phantom will be extremely important to understand how to extract all useful information from MRI for radiomics and AI analysis schemes.

Questions and comments:

- How stable is hydrogel?
    - Hydrogel is very stable for years. It are much more stable than agarose. The actual structure is stable even without water. It forms very robust polymer network. If you can bind the doping in the polymer, it is extremely stable.
- Clustering iron is added to hydrogel. How stable is it when you start to disturb the hydrogel structure by introducing cluster dopes?
    - We hope the cluster dopes will change relaxation time and change diffusion. They are used to modify some parameters. We will need to understand the stability.

<u>QIN Benchmarks, Challengers, and Collaborative Projects update</u>

Dr. Farahani, NCI, presented an update on QIN research. QIN has approximately 25 teams across the US. Many of the network-wide activities are pursued in the context of Challenges and Collaborative Projects (CCPs). The CCPs cover a wide range of research from basic validation concepts to clinical translation of algorithms or tools that have direct relevance to quantitative imaging. He then briefly reviewed each work areas.

ECOG-ACRIN datasets for QIN use: ECOG-ACRIN has contributed data from 14 clinical trials to TCIA. These data will be initially sequestered for use in QIN CCPs and other QIN projects. These data will be made available on TCIA with one year embargo for QIN use.

Current QIN collaborative projects: There are currently six ongoing collaborative projects, i.e., CBV-DSC MR, Prostate DWI MRI, PET hypoxia phantom study; BIDS pipeline; MRI biomarkers for tumor hypoxia prediction; CT feature ontology. Some of these are close to completion.

Challenges: QIN is organizing two challenges, namely Crowds Cure Cancer Annotation and ISBI 2018 Lung Nodule Malignancy Precision. The first uses ECOG-ACRIN trial data for annotation of tumors, and the second uses two time points from LDCT data to predict malignancy.

QIN benchmarks: QIN formed a task force composed of members from CIP, QIN program staff, and representatives from each of the WGs within QIN. They proposed a set benchmarks to define the development stages of QI software tools. A consensus framework for benchmarking QIN tools has five levels, i.e., pre-benchmark, basic benchmark, technical test benchmark, clinical trial benchmark, and clinical use benchmark. The process starts when a PI submits an application with supporting materials, which then is discussed on the coordinating committee monthly meetings for a determination of benchmark level. This process is very valuable in staging QI products, producing standard labels for development and translation of tools, guiding cataloging the QI products, and gauging tools for clinical trials. Many of the QIN products have been qualified at varying levels of benchmarks.

CBIIT Imaging Data Commons: This is a project under CBIIT, and is part of the larger NCI Cancer Research Data Commons (CRDC). QIN is helping manage this program which is about to launch this summer. CRDC was created on the Cancer Moonshot Blue Ribbon Panel recommendations of enhancing data sharing by building a national data ecosystem. The overall goal is to enable all participants across the cancer research and care continuum to contribute, access, combine and analyze diverse data that will enable new discoveries and lead to lowering the burden of cancer. He described the structure of the CRDC which consists of domain-specific repositories (e.g., genomic, proteomic, and imaging), and the status of the various components.

Two Announcements for upcoming activities: Medical Physics Special Issue: Dataset Articles from the Cancer Imaging Archive is especially interested in submissions from contributors who would like to publish their analyses dataset derived from TCIA image datasets. It is required that all datasets be submitted and published on the TCIA website prior to manuscript submission. The deadline for manuscript submission is December 20, 2019. The second announcement is that QIN is continuing collaboration with MICCAI on the BraTS multimodal Brain Tumor Segmentation Challenge 2019 and Computational Precisions Medicine 2019 on mpMRI and Digital Pathology Challenge on brain tumor classification. New this year is that QIN is implementing the challenges on NCI Cloud, which will open next month.

Associate Member Presentation

Dr. Pushpa Tandon, NCI, introduced another category of the QIN membership–associated members. These are investigators who are interested in and work in the field of QI. They will need to submit an application, and if approved by QIN they will be accepted in the program as an associated member. These investigators are not funded by NCI. They enjoy the benefit of participating in annual meetings, working groups and challenges. Many of them have developed collaborations with QIN members. QIN currently has 24 associated members from 11 countries. Dr. Tandon encouraged interested investigators to contact QIN and join as an associated member.

Getting SI traceability of MRI-biomarker measurements

Dr. Stephen Russek, NIST, spoke on NIST's effort on building NIST MRI standards to get SI traceability (i.e., the ground truth) into MRI-biomarker measurements to achieve confidence in measurements precision. He explained how the NIST reference, i.e., SI-traceable ground truth, is used to quantify the measurement uncertainty (variance and bias) in a recent round-robin T1 quantification using data collected from 21 scanners in 9 sites. He showed a plot of variable flip angle (VFA) T1 deviation across the 14 elements in the phantom. The NIST reference value serves as the SI-traceable ground truth with a typical uncertainty of ±1%, whereas the clinical values have a ±20-40% uncertainty. This phantom also helps to understand what part of the scanning protocol leads to the large deviation from the ground truth.

He then explained how they obtain NIST reference values and why they are important to provide a path for extending traceability into human body. NIST provides a calibrated traceable phantom which enables tracing the NMR property measurements (e.g., T1, T2 and ADC) from the human body back to the units of measurements in the primary calibrations of the phantom. The NMR measurement uncertainty is determined by T1 test-retest data to show measurement bias and variance using the NIST Ni-S25 standard (NIST SRM 3136 Ni^{++} in water). The values of bias and variance for NMR T1/T2 are calculated by Monte Carlo simulations of the entire system (i.e., the full model) incorporating all variables. The relative magnetic susceptibility can also be measured and traceable.

He called attention to the NIST website for information on NIST MRI biomarker measurement services including a few calibration services that issue calibration certificates. NIST also has a phantom lending library from which users can lease SI traceable phantoms at low cost. The phantoms are maintained and verified for stability by NIST, users can correlate images on the same object among sites and access common databases/software while assuming no liability in case that a phantom is damaged while in their possession. These phantoms can be shipped within and outside the US.

Dr. Russek lastly pointed out that NIST QI started in 2007, at the same time when NCI QIN and RSNA/QIBA were initiated. He acknowledged the generous support NIST QI received from Dr. Dan Sullivan of QIBA and from Dr. Larry Clarke of QIN.

Questions and comments:

- This type of mono-exponential behavior phantom can be very useful. However, even with high resolution MRI, tissues are non-homogeneous and there are more than one tissue compartment. Do you plan to build phantoms to address higher exponential properties?
    - We are developing tissue mimics for biomimetric phantoms, which will have either biexponential or more complex behavior. There are others in the field also building more complex phantoms. The NIST lending library is happy to receive others' phantoms and to be the distribution network.

<u>Imaging Workflows and Normal Data for Standard Functional Scintigraphic Imaging Procedures in Mice</u>

Dr. Winfried Brenner and his research group at the small animal imaging core facility at Charite realized a few years ago that there were a lack of preclinical workflow and normal organ data for standard scintigraphic imaging in mice. In contrast to clinical functional imaging which has well

established imaging procedures and normal organ values, for example, in quantitative uptake and excretion patterns in kidney and thyroid, there was hardly any scintigraphic data available for normal organ function in mice, nor established imaging procedures.

They set out to study the functional organ imaging (microSPECT/CT and microPET-MRI) in small animals to establish workflows for various tracers and to understand the factors that affect tracers' biodistribution and kinetics to inform the choice of radiotracers that are relevant in clinical settings. He used examples of brain (99mTc-ECD and 99mTc-HMPAO) and bone (99mTc-MDP, 99mTc-HPD and 99mTc-DPD) tracers to demonstrate that some tracers may have very different biodistribution and pharmacokinetics in humans than in animals. They also observed significant effects of sex, age and circadian rhythm on normal organ update and tracer kinetics, such as thyroid, parotid gland and submandibular-sublingual salivary gland complex uptake of 99mTc-pertechnetate and renal function measured with 99mTc-mercaptoacetyltriglycine. In each case, the imaging protocol and workflow was decided based on tracer properties and the organ of interest.

He also outlined future needs for preclinical imaging, including the need for standardization of tracers and workflows for scintigraphic imaging procedures in mice (and rats) to understand which tracers work for which workflows, the need for normal data sets with respect to sex, age, chronobiology and animal species, the need for reduction of variance and thus number of animals for statistical reasons, the need for animal protection, the need for improvement of comparability of studies and data, and the need of an open access database for animal study workflows and normal data which QIN or CIRP may consider to support.

Questions and comments:

- Thyroid uptake is highly dependent on age and sex in mice, but there isn't evidence of such dependency in patients with hypo- or hyper-thyroidism and cancer. They are working with endocrinologists to understand if it is feasible to study this in normal human subjects.

Working Group Breakouts

The PET/CT WG sessions on Day 1 and 2 were attended by:

- Dr. Sandy Napel, Stanford
- Dr. Reichard Beichel, Univ Iowa
- Dr. Ivan Yeung, Princess Margret Cancer Center, Toronto
- Dr. Carlos Uribe, Univ. of British Columbia Cancer Center, Vancouver
- Dr. Binsheng Zhao, Columbia Univ.
- Dr. Paul Kinahan, Univ. of Washington
- Dr. Heang-Ping Chan, Univ. of Michigan
- Dr. Hadjiyski, Univ. of Michigan

Future projects discussed were development of a 3-D printed phantom with variable density/shape "lesion" insert components that could be used for CT scanner harmonization and better quantitation of tumor heterogeneity and radiomics features. If engineered for PET-filling the phantom could be used to calibrate TOF reconstruction.

Dr. Kinahan affirmed the value of this kind of phantom calibration work pointing out the development of a phantom with texture features developed by Scott Wollenweber at GE, studies with certain fluid-filled kidney phantoms, and some of his own work with a DRO used in the 18FLT breast cancer challenge and other PET/CT investigations. Others pointed out that a 3-D printed phantom could utilize some of the simulated lesion data that Duke workers has inserted into actual scan file images to assess CT radiomics features. Pulling radiomics features out of PET images would be more problematic given the lower resolution of this modality although some progress has been made in categorizing FDG-PET lesions using semantic features.

The CTDD WG sessions on Day 1 and 2 were attended by:

The CTDD WG will consult with the Clinical Consulting Committee who reviewed and evaluated the QI tools during the meeting to understand what tools are shovel ready for what levels of clinical evaluation, and plan out the details on their activities for the next year accordingly.

The WG proposed several actionable next steps for the coming year.

- Publish a white paper on standardization (target journal JNM)
- Understand QI tools for potential testing in NCI Cooperative Groups, and understand which technically validated QI tools can be matched to Cooperative Groups' trials for clinical assessment.
- Prepare a Consumer Reports style document for each of the QI tools to summarize the key considerations for clinical assessment.
    - criteria of benchmarking
    - context of diseases
    - preferred disease sites
    - tools scores ≥ 3 on the QIN benchmarking scale
    - ECTCN vs. NCTN
    - Additional burden of adding this QIN tool to trials
- Increase the awareness and visibility of QIN tools and approach Cooperative Groups' imaging chairs
    - Post technically validated QI tools on the QIN website
    - Create a focused list of tools for NCI Cooperative Groups' imaging chairs to review
    - Identify imaging champions in disease committees
    - Prepare presentations to Alliance about QIN tools
    - Understand the imaging needs:
        - IROC to prepare a list of imaging needs in a questionnaire
        - QIN investigators to come to IROC semiannual meeting and be exposed to clinical trial concepts to understand imaging needs