Wright Center of Innovation in Biomedical Imaging
The Ohio State University
Wexner Medical Center

IROC Imaging and Radiation Oncology Core
NCI National Clinical Trials Network

The Impact of QC in Multicenter Clinical Trials - The IROC Experience for NCTN Focusing on ^{18}F-FDG-PET

David Poon, Preethi Subramanian, Richard Jacko, Prayna Bhatia, Tim Sbory, Shivangi Vora, Ajay Siva, Talha Saif, Jun Zhang, Michael V. Knopp

Affiliation:
- Wright Center of Innovation in Biomedical Imaging, Department of Radiology, The Ohio State University, Columbus, OH

Acknowledgements:
- The funding of IROC Ohio through the NCI grant 5U24CA180803 is gratefully acknowledged and well as the infrastructure support by the Ohio Third Frontier TECH 09-028.

Contact:
- Michael V. Knopp, MD, PhD (PI)
  E-Mail: knopp.16@osu.edu
- Jun Zhang, PhD, DABR
  E-Mail: zhang@IROCOhio.org

RSNA 2016

Quality Assurance: A Reality Check?

In multi-center clinical trial environments, there are many factors influencing the acquisition of images including, differing local standard of care imaging operating procedures, range of scanner models, technical operator variance among many other factors. Quality assurance is critical to acquire robust imaging data while using standardized protocol implementations that enable a broad range of clinical site and imaging services to participate.

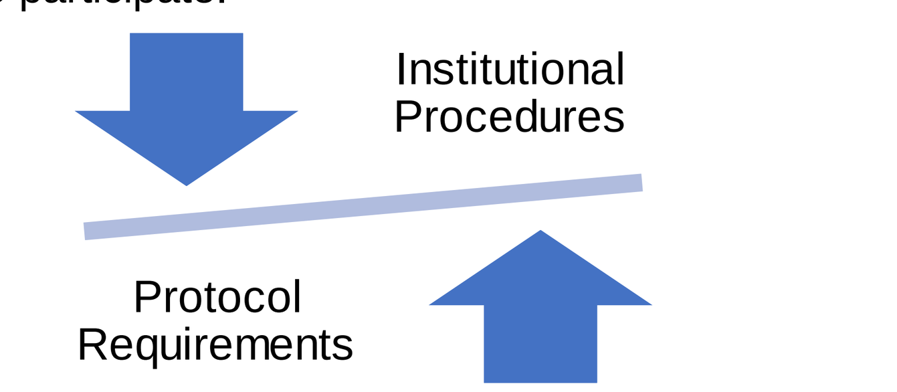

Figure 1: PET/CT imaging within clinical trials of the NCTN are frequently performed as Standard of Care imaging, which means that they are coverable by insurance carriers. While is fully appropriate, it leads to challenges when institutional procedures and clinical trial protocol requirements are conflicting. If those lead to imaging protocol variations, they can be readily detected by advanced QC approaches and most often resolved by education, training and institutional engagement.

Clinical trials with an ^{18}F-FDG-PET imaging component depend on consistent imaging acquisitions for clinical response assessments for treatment decisions. A DICOM based assessment of submitted images, allows local imaging parameters to be compared with protocol specifications. Due to the reality of varying local imaging practices, it is important to help standardize acquisition and patient preparation techniques. While phantom based calibration for initial site credentialing establishes a solid foundation for participation in a multi-center clinical trial, individual patient imaging should be monitored for adherence of the research standards expected by the trial committee.

Imaging and Radiation Oncology Core (IROC)

The Imaging and Radiation Oncology Core (IROC) cooperative was formed with the reorganization of the National Cancer Institute's National Clinical Trial Network (NCTN) and started to provide network wide services since March 2014. IROC Ohio is one of six imaging core laboratories within the cooperative and focuses on supporting and managing NCTN trials for the Alliance and SWOG network groups.

Our broad spectrum of services, a selection of which is listed below, puts us in prime position to analyze, educate and standardize acquisition even in a multi-center clinical trial environment.

- Protocol development support
- Site credentialing (equipment validation, test patient data assessment)
- Site personnel training & education
- Data quality assurance, banking and case management.
- Real time as well as end point data analysis and review.

Getting the most out of DICOM Metadata

A wealth of information pertaining to image acquisition can be obtained from the DICOM metadata for each examination that is stored in each image file. The metadata is composed of DICOM Tags assigned to a broad range of parameters, classified either as public or private according to the vendor's conformance statement.

We have identified the DICOM tags linked to the key parameters that reflect the quality and consistency of a ^{18}F-FDG PET/CT acquisition as detailed in **Table 1**. Thus, these tags can collectively act as a blueprint to analyze any given examination to assess consistency with a specific clinical trial protocol and deviation that maybe due to a imaging facility's unique standard operating procedures.

Analyzing the DICOM tags is a significant quality assurance tool as it enables an objective and detailed assessment of the performed imaging scans. Inconsistencies with respect to both acquisition as well as system calibration information can be determined. DICOM metadata are thus key enablers of examination based quality assurance essential in multi-center clinical trials.

Table 1: This table describes a several DICOM tags pertinent to ^{18}F-FDG PET/CT examinations along with the standard descriptions. Taken together with other tags, a check of key DICOM information can characterize a local imaging facility's procedures of ^{18}F-FDG PET/CT image acquisition and protocol adherence.

General Parameters	
Vendor Make	0008, 0070
Vendor Model	0008, 1090
Software Version	0010, 1181
PET Specific Parameters	
Patient Position	0018, 5100
Radiopharmaceutical Injection Time	0018, 1072
Radiopharmaceutical Total Dose	0018, 1074
Image Acqusition Time	0008, 0032
Reconstruction Method	0054, 1103

Methodology

We evaluate every imaging study submitted not only in terms of visual quality but also, using a DICOM based assessment to check for protocol compliance. A semi-automated program was developed to extract imaging parameters from the DICOM headers which are then recorded (**Figure 2**). This matrix (**A**) was then used to established a heat-mapping range of parameters (**C**) based on their impact on imaging quality & quantification, considering the protocol expectations & local practices (**B**). This heat-mapping range is further classified into: i) in complete agreement with the protocol (***Green***), ii) out of range while still acceptable (***Yellow***), or iii) both out of range and not acceptable for evaluations (***Red***). Using this heat-map, a weighted score based semi-automated QC approach (**D**) is used to classify the submitted imaging exam into one of the three categories (**E**). A detailed QC report and communication with sites in a robust feedback mechanism ensures consistent imaging acquisition and protocol adherence (**F**).

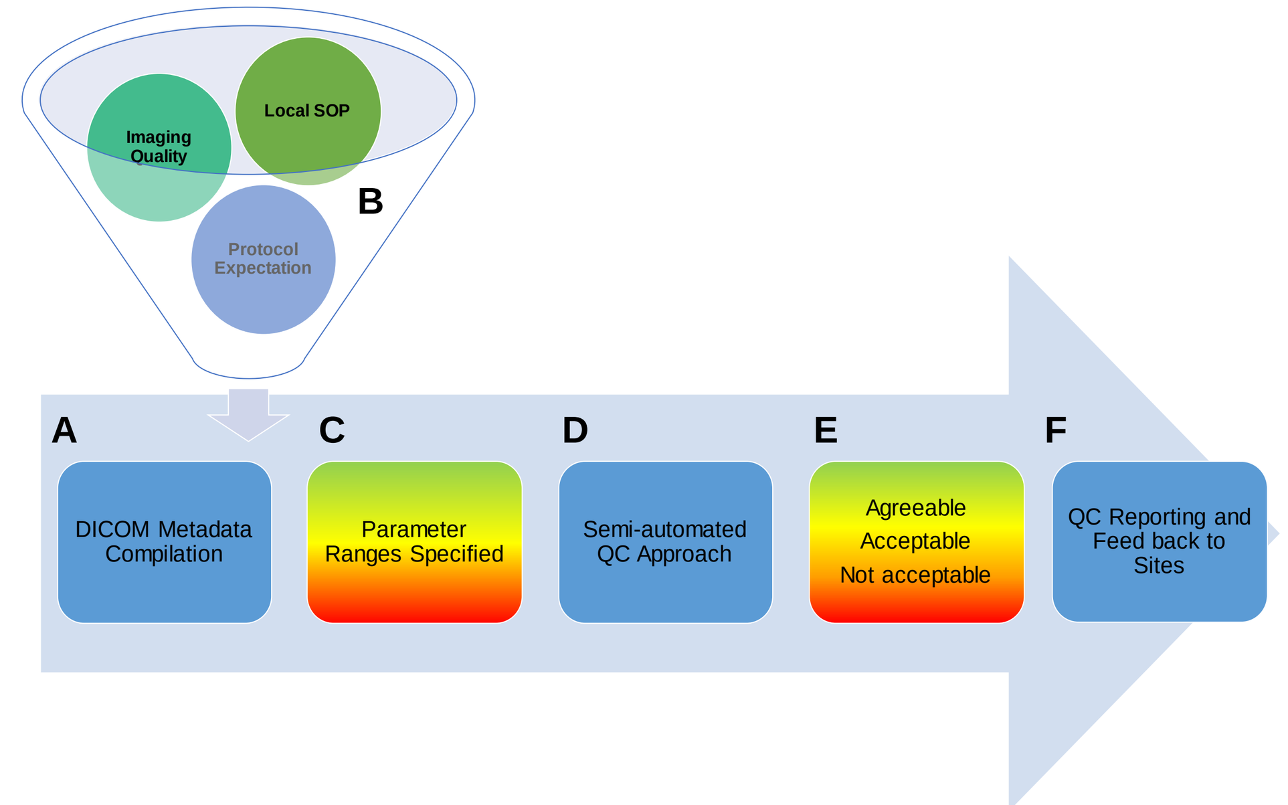

Figure 2: This figure highlights the workflow of IROC quality assurance methodology. DICOM Metadata compilation (**A**) when analyzed through input from the protocol imaging committee, and considering the local SOP as well as its effect on imaging quality (**B**), we were able to come with a parameter heat-mapping distribution (**C**). This then allows us to use a semi-automated approach (**D**) to assess images using a weighted parametric approach (**E**). Communication with sites in a feedback loop mechanism (**F**) has resulted in an increased rate of imaging compliance over the lifetime of the clinical trial.

QC Methodology: ^{18}F-FDG-PET/CT Study Weighted Scoring

To better characterize the overall compliance of a FDG-PET/CT study, a weighted scoring system was developed. This was achieved through communication with the trial imaging committee to discuss the most important aspects of a FDG-PET/CT study in terms of data reproducibility and how they proposed the data should be analyzed upon the trial's completion. A total of fifteen key factors are considered (**Table 2**). These factors range from when the exam was acquired, while being on study treatment, to DICOM specific factors relating to the imaging acquisition.

If a factor is found to deviate from the protocol guidelines, it is given a 'disagreeing but acceptable' (yellow) or 'not acceptable' (red) score based on the severity of the deviation. How these severities are defined as 'acceptable' vs 'not acceptable' were established by the parameter ranges based on imaging committee input. These factors and ranges can be protocol specific, allowing this methodology to apply to a range of clinical trials and imaging modalities.

Table 2: A multi-factor heat-mapping system developed for a multi-center ^{18}F-FDG-PET/CT trial. Of the factors considered, factors 1, 2, 3, 5, and 6 are assigned a more heavily contributing value, as they were considered more important to the viability of the ^{18}F-FDG-PET/CT scan in terms of assessing a consistent treatment response (factor 1), to the ability to calculate SUV (factor 5) and standardizing SUV calculations (factors 2, 3, 6). Green, Yellow, or Red categories influence a factor's given score in terms of being in agreement to the protocol, disagreeing but acceptable, or not acceptable, respectively. Collectively, these scores add up to a possible total of 100, characterizing an entire exam.

Factor #	QC Items	Weighted Score	Green	Score	Yellow	Score	Red	Score
1	Scan window of PET/CT Exam	10	Within specified window	10	< 7 days difference	6	>7days	1
2	Emission uptake time (minutes)	15	50-70	15	45-90	8	<45 or >90	0
3	Consistency of Uptake Time (Baseline/Follow-up)	15	<10 minute difference	15	< 15 minute difference	8	>15	0
4	Completeness of imaging data	3	Complete	3	Includes Key Sequences	2	Incomplete	0
5	Data imaging format	10	DICOM	10	Secondary-DICOM	3	Non-DICOM	0
6	Consistency of PET/CT Scanner	15	Same	15	N/A	0	Different	5
7	Consistency of arm positioning	5	Same	5	Different	3	N/A	0
8	Consistency of scan direction	5	Same	5	Different	3	N/A	0
9	Injection site	3	No Extravasation	3	Clear Extravasation	1	N/A	0
10	De-identification	2	De-identified	2	Not de-identified	1	N/A	0
11	Image resolution	3	Meets protocol	3	1mm difference	1	>1mm difference	0
12	Case Report Forms	2	Complete	2	Incomplete	1	N/A	0
13	Fasting period (hours)	2	>4	2	N/A	0	<4	0
14	Glucose level (mg/dL)	5	<200	5	N/A	0	≥ 200	1
15	FDG Dosage (mCi)	5	8-20	5	7-22	4	<7 or >22	1

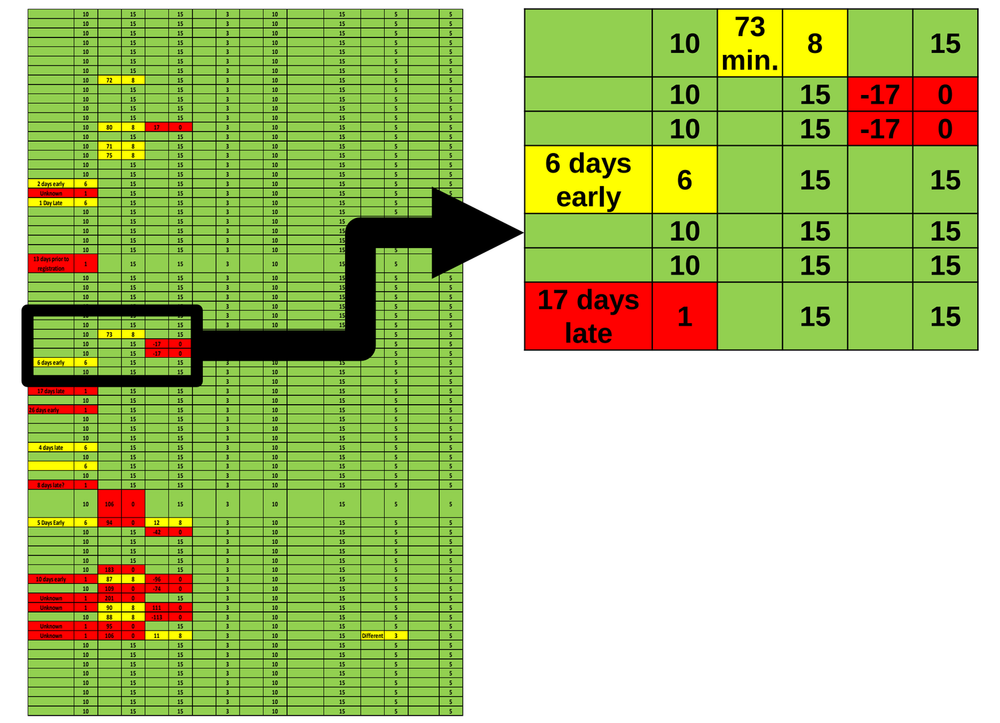

Table 3: A sample selection of a semi-automated heat-map matrix of several individual ^{18}F-FDG PET/CT exams submitted for a multi-center clinical trial. Each line represents one exam, with each factor considered sequentially from left to right. The analysis is color-coded for each factor level. By setting filters, specific criteria can be mapped and factors outside of the optimal protocol ranges (yellow or red) identified. Each individual factor's score, categorized by green, yellow, or red, is added up to a total out of a maximum of 100. The total score of an individual exam can then be further categorized into green, yellow, and red, marking an exam as overall in protocol agreement, acceptable with minor deviations from the protocol, or unacceptable with major protocol deviations.

The categorization of several exam factors by their green/yellow/red scores can be arranged into a heat map matrix to assess their adherence to protocol guidelines. This matrix incorporates factors of all submitted studies for a clinical trial. This matrix can be zoomed in and out to pinpoint 'hotspots' i.e. factors that may not conform to optimal parameter range. In this way feedback can be tailored to sites in order to provide the most affective means of educating best practices (**Table 3**).

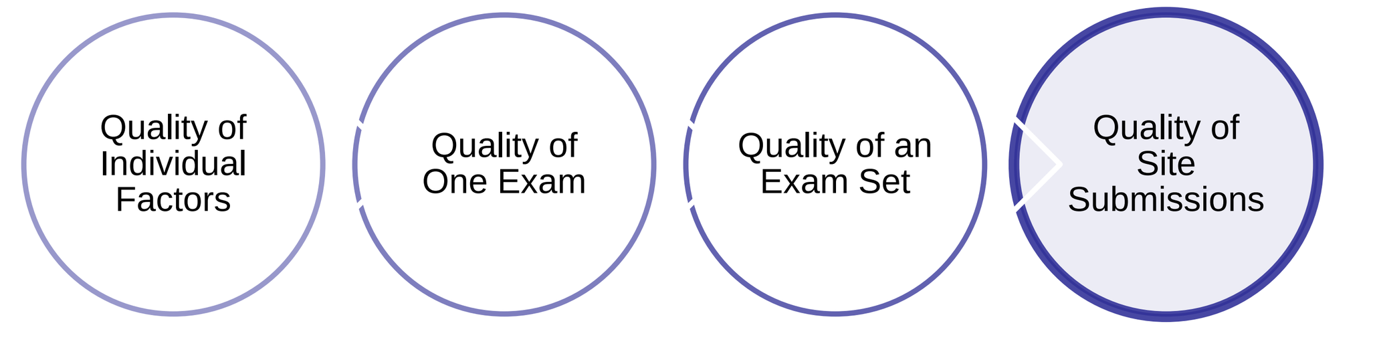

Figure 3: Heat-map based quality assurance to examine the quality of submitted data allows quality levels to be examined from individual parameters to the overall quality of imaging site submissions.

Since each ^{18}F-FDG PET/CT exam is comprised of several factors, not only is each individual factor's score considered, their weighted sum of scores (out of 100) as a whole categorizes a single exam's overall agreement with protocol guidelines. Factors with scores that are weighted more heavily have a greater impact on this total, if found to be disagreeing with the optimal parameters. In studies that involve the submission of multiple exams for a single subject over the course of a trial, individual score totals can be analyzed to assess site imaging capabilities. Therefore the quality assurance mechanism is able to assess imaging submissions over several orders of magnitude (**Figure 3**).

SUV: A Fortress Built on Sand?

Standard Uptake Value (SUV) based response assessments depend on the premise of consistent scanning techniques and patient preparation, in order to gauge the true treatment response. This is especially important when a clinical study uses SUV to make treatment decisions based on a just-in-time central imaging (adaptive) review, or when SUV is a component of an end-point analysis. Several factors can affect the consistency of an SUV calculation, from the time a patient is injected with the radiopharmaceutical to when the PET emission begins, to the period of time a patient fasts prior to the exam, among others. Therefore, it is critical to eliminate such external factors for these calculations to be a truly reliable and thus consistency needs to be assess and confirmed.

Another key aspect is the consistency of these parameters with respect to the initial exam (baseline), as a patient is being monitored for treatment effects over a long period of time. Data anonymization is a powerful tool, as the SUV calculation may not be accurate if certain DICOM metadata cannot be found or have been replaced with dummy values (**Figure 4**).

Figure 5: PET/CT built from sand as seen at the Brazilian SNM meeting in Rio (2015). We use this analogy to highlight that SUV depends on many factors and when well managed enables an essential semi-quantitative readout.

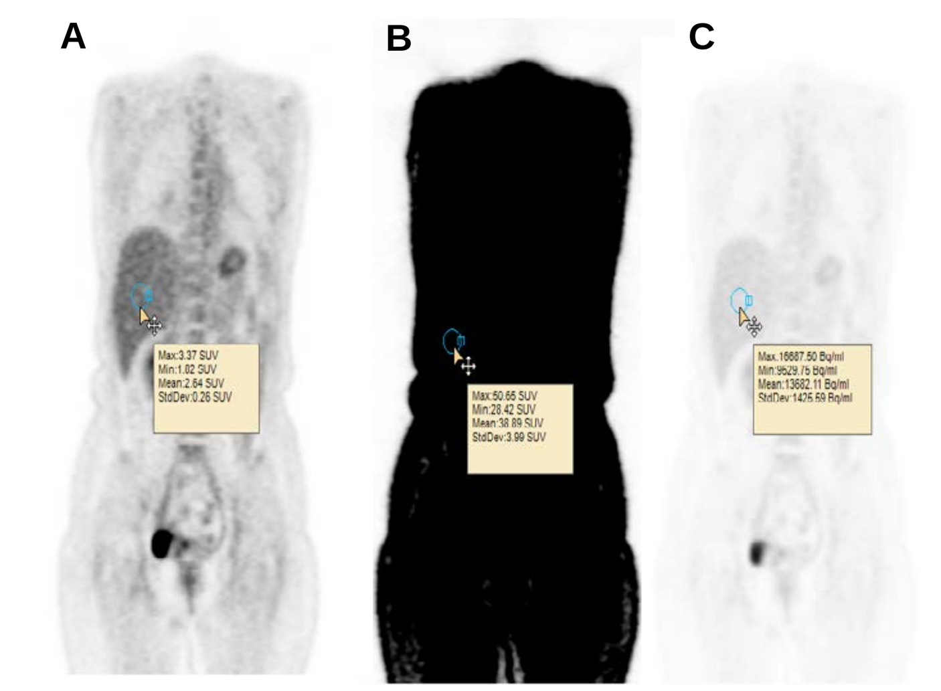

Figure 4: An comparison of ^{18}F-FDG PET examinations with inaccurate SUV displays. Several factors can result in an erroneous or un-evaluable SUV measurement for ^{18}F-FDG PET exams, such as non-compliant PET emission uptake or an inappropriate fasting period. Some of these factors are not a result of out-of-range parameters, but of instances of data anonymization. SUV calculation is computed when a patient's weight (**A**) at the time of acquisition is curated in DICOM. It is not uncommon for sites' anonymization procedures to alter this value with either a dummy value (1000 Kg) (**B**), or deleted (**C**). This results in erroneous computation of SUV (**B**) or the calculation is ignored altogether and replaced by the radiopharmaceutical activity measurement being displayed (**C**).

Even though several of these parameters have to do with patient preparation and not necessarily with the acquisition of images, such critical information is recorded in the examination's DICOM image tags. Factors influencing SUV (i.e. patient preparation, etc.) that cannot be pulled from the DICOM metadata can be clarified via proactive site communication. There are many potential sources that can impact the quantitative assessment thus necessitating a robust QC approach.

Identifying the Impact of Compliance in ^{18}F-FDG-PET/CT

Since several factors of an ^{18}F-FDG PET/CT exam can be analyzed via the heat-mapping approach compliance trends focusing on individual parameters that can affect the quality of an exam can be readily followed. We have been able to observe how communicating quality assurance findings and feedback to sites can affect the quality of an exam, as sites improve their submitted imaging data, the overall submitted exam down to their individual parameters (**Figure 5**).

A primary goal of the developed heat-mapping mechanism is to avoid categorizing submitted imaging exams into only an acceptable vs not-acceptable format. Whether observing trends in the overall study conformance to acquisition guidelines to pinpointing individual parameters, we are able to observe the general quality trends of submitted ^{18}F-FDG PET/CT studies over the lifetime of a clinical trial. A heat-based mapping of protocol compliance at a varying level as summarized in **Figure 3** enables a more meaningful insight into the true clinical trial performance than even a more rigid, cut-and-dry assessment of imaging protocol conformance or non-conformance.

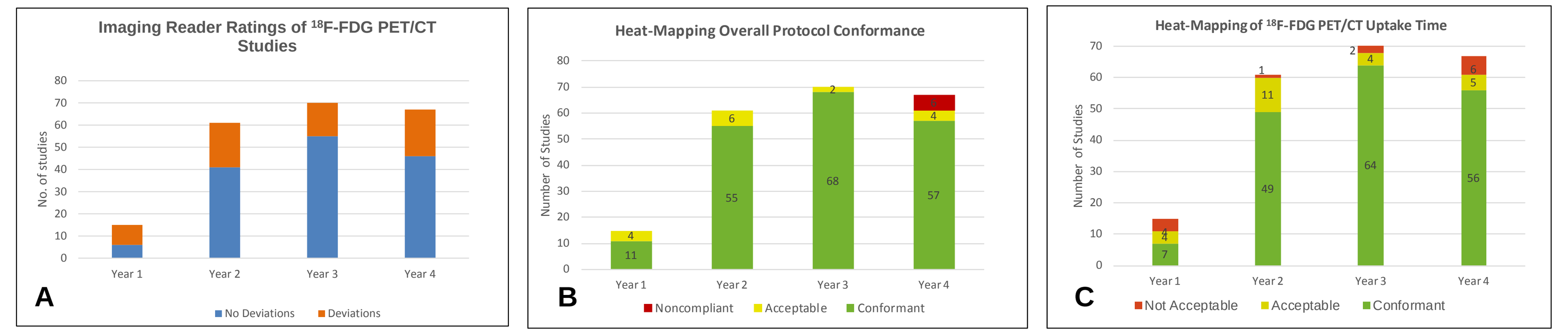
Imaging Reader Ratings of 18F-FDG PET/CT Studies
Heat-Mapping Overall Protocol Conformance
Heat-Mapping of 18F-FDG PET/CT Uptake Time
A
B
C
Year 1
Year 2
Year 3
Year 4
No Deviations
Deviations
Nonconformant
Acceptable
Conformant
Not Acceptable

Figure 6: A comparison of 213 PET examinations within a NCTN clinical trial over a four year period on which we performed QC found deviations to protocol recommendations vs those that were found to conform (no deviations). (**A**) When asking the blinded reader to assess during their read if the image acquisition was performed according to protocol a large subset (orange) was deemed to have deviations; (**B**) When using a rigorous, consistent rule based applied heat-map approach, the overall compliance was found to be substantially better as shown. An additional advantage of the heat-mapping mechanism is that individual quality factors can be identified for protocol conformity; (**C**) Details the compliance in regard to the trial protocol defined uptake time which can be readily visualized using the heat-map approach. When a site used a deviated uptake time, but did do it consistent for the patient, the patient studies were still evaluable for the purpose of assess response and therefore were overall categorized as acceptable. Though trends toward a greater proportion of compliance can be seen in all three graphs, simply categorizing a study as having deviations/no deviations to study directed parameters does not fully make clear an exam's usefulness for real-time or end-point analysis.

Over the course of four year accruing multi-center ^{18}FDG-PET/CT trial, implementing quality-control measures managed to achieve an increase in conformant imaging study parameters from 73% conformance at the end of year 1 to 85% conformance by the end of year 4 (**Figure 6B**). If we review on an individual parameter essential to be kept consistent across multiple studies within a single patient, the PET emission imaging post injection uptake time, we are also able to observe an increase of conformant PET emission uptake from 47% to 84% (**Figure 6C**).

While observing broadly that there are is a percentage of studies over the course of the first four years of the trial that deviated from the acquisition guidelines, these percentages do not reflect accurately whether or not the non-conformant exams are useful for data analysis.

With all factors of quality considered through heat-mapping, the number of studies able to be analyzed is not significantly impacted: while 68% of exams acquired at year 4 are technically non-conformant to the acquisition guidelines (**Figure 6A**), 90% of the exams collected that year were able to be utilized for the study end-point analysis, and throughout years 1 through 3, all studies, regardless of whether or not they contained a deficient quality factor, had parameters that collectively identified the studies as evaluable for data analysis when interpreted through the heat-mapping mechanism (**Figure 6B**).

Summary & Conclusion

- As more institutions become interested in enrolling patients to clinical trials where quality imaging is essential for therapy assessment and disease staging, a continuing relationship with QA centers and sites can help to ensure that the highest quality and validity of data can be achieved.
- With a multi-factor based approach to identify acceptable vs not-acceptable submitted PET/CT data, a quality driven QC methodology has shown to achieve substantial improvements in all aspects of trial performance.
- By obtaining quality-related information from the imaging exam DICOM header, a reliable method is established for quality assurance that then can build on a database of study-specific acceptable parameters that allows a more practical and meaningful categorization and enables improved clinical trial performance.
- Heat-mapping of QC results allows for the observance of trends and areas of quality that may most impact trial data and can be readily used a QC trial management and quality improvement efforts.
- QC for clinical PET/CT trials has expanded from the traditional phantom based device qualification to a patient and exam based approach that is essential to secure data quality and scientific precision while enabling the broader community to participate in the essential NCTN trials.
- The introduced quality management approach assists greatly in identifying major issues and creating a feedback loop to participating institutions with a positive impact on data quality of the imaging data generated.

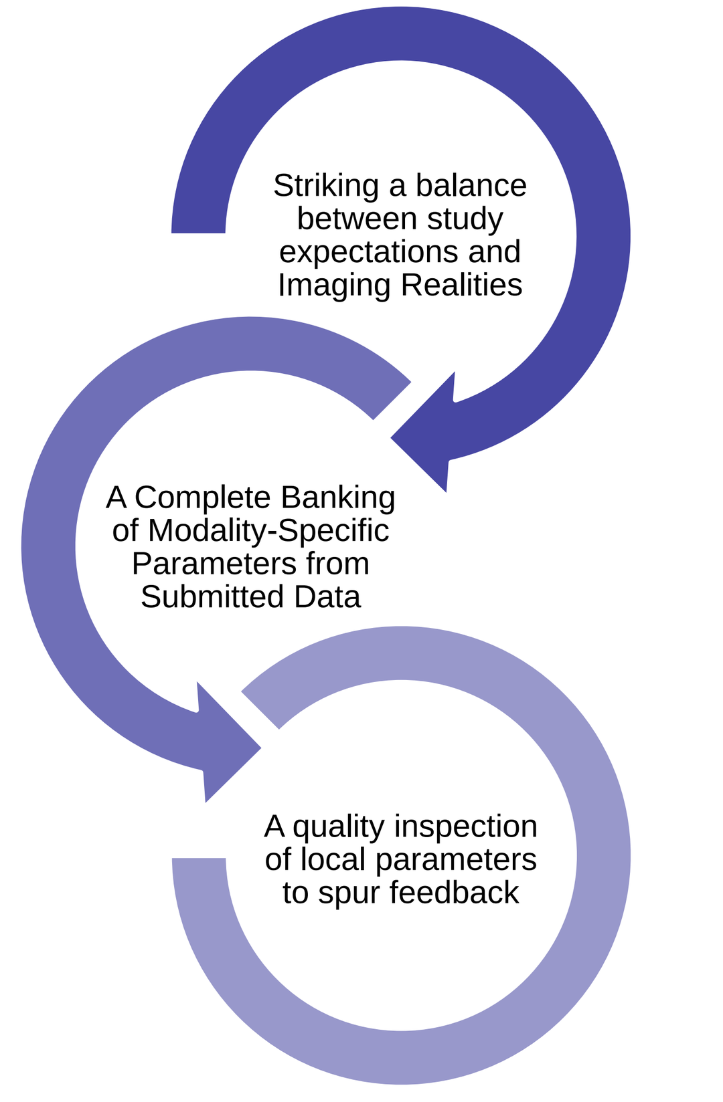

Figure 7: Quality assurance of clinical trials performed within standard of care while requiring protocol specific characteristics requires innovative approaches that require a balance with clinical reality. Advanced developed QC tools like heat-mapping for proactive communication improve overall quality of clinical trials.