FDG-PET/CT UPICT V 2.0
Full document, publicly reviewed version (Dec 2014)

Table of Contents

Major Headers, UPICT V2.0, QIBA FDG-PET/ CT Imaging Protocol

X. Imaging Protocol

Executive Summary

The FDG-PET/CT subgroup of the Uniform Protocols for Imaging in Clinical Trials (UPICT) Working Group (now part of QIBA initiative), consisting of imaging physicians and medical physicists worldwide with expertise in early drug development from academic research organizations, government and industry, together with imaging specialists, has met regularly through in-person meetings and weekly conference calls over the last 5 years to develop these evidence-based consensus guidelines for the use of FDG-PET/CT in oncology clinical trials. A critical component of the development process was to extract 'verbatim' information from acknowledged key scientific publications on FDG-PET in clinical trials (references) into the appropriate section of the UPICT template; consolidate the information and from the consolidated material, develop consensus statements (where appropriate), identify gaps in scientific knowledge and suggest areas where future investigation may be warranted. In particular, society guidance documents from EORTC [1], NCI[2], and EANM [3] as well as Standard Operating Procedures from ACRIN imaging core laboratory [4, 5, 6, 7, 8] were used as references. The process of conversion from consolidated to consensus was accomplished by the UPICT group in conjunction with input from the SNM FDG-PET Global Harmonization Summit held in Salt Lake City in 2010.

This UPICT Protocol is intended to guide the performance of whole-body FDG-PET/CT within the context of single- and multi-center clinical trials of oncologic therapies by providing *acceptable (minimum), target, and ideal standards for all phases of the imaging examination as defined by the UPICT Template V1.0,* with the aim of minimizing intra- and inter-subject, intra- and inter-platform, inter-examination, and inter-institutional variability of primary and/or derived data that might be attributable to factors other than the index intervention under investigation.

This protocol is complementary to the recently published *FDG PET/CT: EANM Procedure Guidelines for Tumour PET/CT Imaging: version 2.0* [9], which is primarily focused on the clinical and quantitative FDG PET/CT imaging procedure. There is some inevitable overlap in the two documents but the emphases and goals are different. While focused primarily on the use of FDG-PET/CT in the conduct of oncologic clinical trials, the UPICT protocol has significant utility in reducing the bias and variability of imaging studies in clinical practice. Similarly, the *EANM Guidelines* contain information useful in conducting clinical trials, but not in the depth provided in the UPICT protocol.

The specific potential utilities for the FDG-PET/CT study(ies) as performed in accordance with this Protocol within any particular clinical trial could be to utilize qualitative, semi-quantitative, and/or quantitative data for single time point assessments (e.g. diagnosis, staging, eligibility assessment, investigation of predictive and/or prognostic biomarker(s)) and/or for multi-time point comparative assessments (e.g. response assessment, investigation of predictive and/or prognostic biomarker(s)). More generally, such standardization of FDG-PET/CT within the conduct of clinical trials should:

1) Support internal decision-making in drug, biologic, and device development,
2) Provide data to support registration and market-label indications, and
3) Support the qualification of FDG-PET as an imaging biomarker (including as a surrogate for clinical endpoints) by supporting meta-analyses of multiple clinical trials.

This document includes specifications for the performance of CT for the purposes of attenuation correction and/or localization, but does not address the performance of diagnostic CT within the context of FDG-PET/CT; although the integration of diagnostic CT in conjunction with FDG-PET/CT for oncology is acknowledged as potentially useful and appropriate. When the integration of diagnostic CT is desired as

part of the imaging protocol within the clinical trial, specifications for the CT portion of the imaging protocol may be derived from other UPICT protocol(s).

While focused primarily on the use of FDG-PET/CT in the conduct of oncologic clinical trials, this protocol also may have utility for guiding the performance of high quality imaging studies in clinical practice.

Preamble

The process for developing Uniform Protocols for Imaging in Clinical Trials was formalized with the development of a protocol template during a clinical trial working group session at the CTSA-IRAT meeting, April 29-31, 2009 in Baltimore, MD. An FDG-PET working group was also formed at this meeting, which began regular conference calls to develop a specific UPICT for FDG-PET in oncology. A critical component of the development process was to extract 'verbatim' information from acknowledged key publications on FDG-PET in clinical trials (references) into the appropriate sections of the UPICT template; consolidate the information (with citations) and from the consolidated material, develop consensus statements (where appropriate), identify gaps in scientific knowledge, and suggest areas where future investigation may be warranted. The process of conversion from consolidated to consensus was in part accomplished by the UPICT group, in conjunction with input from the SNM FDG-PET Global Harmonization Summit held in Salt Lake City in Sept 2010.

A bullseye approach to compliance was utilized, as there are practical and technical limitations in the performance that can be achieved at different sites ranging from academic medical centers to community hospitals. Specifically, a minimum "Acceptable" compliance level was established that must be achieved by *all* sites participating in the clinical trial. In addition, a "Target" compliance level is defined to produce higher performance, which is achievable in *some* sites but not required from all sites participating in the clinical trial. Lastly, an "Ideal" standard is defined where the desired characteristics or features are defined, which may not immediately achievable. The "Ideal" standard is often defined to provide guidance to instrumentation and software developers for future product features that would improve quantitative performance in clinical trials. "Exploratory" specifications are provided in areas where there is a current lack of knowledge that may be addressed in the future with the collection of additional data.

Figure 1: Bullseye compliance

Relationship to the QIBA FDG-PET/CT Profile

The publication of this UPICT guideline coincides with the "Profile" technical specifications developed by the FDG-PET/CT Oncology group of the Radiological Society of North America (RSNA) Quantitative Imaging Biomarkers Alliance (QIBA). The QIBA FDG-PET/CT Profile [10] provides both literature-based and consensus recommendations for identified gaps in quantitative imaging. The Profile can be used as a resource for all aspects of quantitative FDG-PET/CT imaging. It is important to note that the protocol components of the Profile were originally based on the 'Acceptable' criteria in the UPICT Protocol and reference the matching sections of the UPICT protocol where appropriate. However, during the writing of the FDG-PET/CT Profile, there have been small changes to the protocol components. Thus while the two documents have similar protocol components, they are not identical.

1. Context of the Imaging Protocol within the Clinical Trial

1.1. Utilities and Endpoints of the Imaging Protocol

The specific utilities for the FDG-PET/CT imaging include:

- diagnosis and staging of tumors [3, 4, 5, 11]
- prognostic stratification / biomarker [5, 11, 12]
- treatment planning or triage [5]
- edge detection of tumors in radiotherapy planning [3]
- lesion localization and characterization [3, 4, 5]
- evaluate and quantify tumor response / predictive stratification / biomarker [2, 3, 6, 7, 11, 12]
- correlation between imaging and tissue biomarkers and/or pathway activity [7]

1.2. Timing of Imaging within the Clinical Trial Calendar

The study protocol should specifically define an acceptable time interval that should separate the performance of FDG-PET/CT image acquisition from both (1) the index intervention and (2) other interventions (e.g. chemotherapy, radiotherapy or prior treatment).

If response assessment will be based on serial FDG-PET/CT imaging studies, the time interval between the baseline study and the initiation of treatment should also be specified as well as the time intervals between subsequent FDG-PET studies and cycles of treatment. *Additionally, the study protocol should specifically define an acceptable timing variance for performance of FDG-PET/CT around each time point at which imaging is specified, i.e., the acceptable window of time during which the imaging may be obtained "on schedule."*

The timing interval and window are entirely dependent upon:

1) the utility for the FDG-PET/CT imaging within the clinical trial
2) the clinical question that is being investigated
3) the specific intervention under investigation

There is some difference of opinion based on the reference source and the specific index intervention. Suggested parameters for timing of FDG-PET/CT within oncologic trials include:

- *When results of FDG-PET/CT are a study entry criterion, the baseline (eligibility) scan(s) ideally should be performed within 21 days before initiation of the therapeutic intervention. It should be noted that tumors with low FDG uptake (also see Sections 9 and 10) may not be suitable for follow-up studies of treatment response with PET [13].*
- *For FDG-avid and evaluable tumors,* the minimum interval between the last dose of chemotherapy *or biologic therapy* and FDG-PET ideally should be 10 days [3], with an acceptable interval of up to 14 days [2, 11].
- As an alternative if FDG-PET/CT is being used during an ongoing treatment schedule (perhaps as an early predictor of response), the test should be performed at an interval within the treatment schedule that is determined by factors including, but not limited to, the type of treatment, specific cancer diagnosis, specific treatment target, and details of the treatment schedule itself. For example, if the FDG-PET/CT will be performed between cycles that have no "break," the scan might be performed as close to the start of the next cycle as possible [3]. *However, if the FDG-PET/CT will be performed within a treatment plan*

that incorporates periodic "breaks" between sets of treatment cycles, the scan might be performed shortly after the completion of the preceding cycle rather than after the "break" and therefore prior to the next cycle.

- In trials of or including radiation treatment, an interval of up to 4 months may be required [11], although many investigators recommend a minimum delay after radiation therapy of 6-8 weeks or longer before performing the post-treatment FDG-PET study [2]. *Studies evaluating completeness of response should be performed later, however investigational studies used to modify therapy or predict outcome may be performed during therapy.*
- When FDG-PET/CT is used for post-treatment response assessment in lymphoma, imaging should not be performed before at least 3 weeks after chemotherapy and preferably 8 – 12 weeks after completion of radiotherapy per the consensus statement of the Imaging Subcommittee of the IHP in Lymphoma [3, 13]. For intra-therapy evaluation please see bullet #3 above.
- *An issue that must be addressed in the study-specific clinical trial protocol is the specific windows about each time point that would constitute an appropriate variance for that specific clinical trial*

1.3. Management of Pre-enrollment Imaging

The imaging protocol must contain documentation as to how pre-enrollment imaging should be managed; specifically 1) whether imaging obtained prior to enrollment is used as baseline imaging, and 2) if so, under what specific conditions.

It is suggested that the specific conditions should take into account technical factors related to the imaging platforms (PET and CT) as well as the biology of the disease and the specific interventions used in the trial. In general, scans performed as standard clinical care on PET/CT scanners that have not been previously qualified for the clinical trial and/or not in conformance with the imaging protocol would not be acceptable for the clinical trial.

One reference suggests that PET/CT scanning performed within eight weeks prior to initiation of drug therapy could be used as the baseline study [6]. While another source states that if the pre-enrollment PET/CT was performed on an imaging platform not approved for use in the trial or otherwise does not meet trial requirements, the scan should be repeated, if feasible within the trial budget; however studies that are performed on approved scanners and otherwise conforming to all trial specification will be accepted as baseline studies and will be subjected to the same QA as studies performed after registration [4, 6].

1.4. Management of Protocol Imaging Performed Off-schedule

Acceptable: The clinical trial protocol should explicitly state the management of FDG-PET/CT (and all other imaging tests) performed on qualified platforms and in accordance with the specifications of the imaging test (see Sections 2.2, and 3 - 7) but outside of the specified time window(s) of scheduled imaging (see Sections 1.2 and 1.3).

The inclusion of data from these off-schedule time points might have significant impact on the data analysis for the clinical trial. Therefore, the study design should state how such off-schedule data points will be managed.

Potential options include, but are not limited to:

1) Using all of these data in addition to the imaging data obtained on-schedule
2) Using only some of these off-schedule data (e.g. FDG-PET/CT obtained as confirmatory to other non-imaging evidence of disease status), in addition to the imaging obtained on-schedule
3) Ignoring all imaging data obtained off-schedule

Unless specifically allowed by the clinical trial protocol, off-schedule imaging should not be allowed to substitute for on-schedule imaging. The clinical protocol, the informed consent document, and the clinical trial budget should address the management of off-schedule imaging that was obtained for clinical purposes in temporal proximity to the necessary on-schedule research imaging.

The clinical trial protocol should also specifically address how off-schedule scans will be managed in the analysis of the clinical trial overall (e.g., will the sample size be inflated to allow for post hoc exclusion of subjects who drop out secondary to findings noted on off-schedule imaging studies).

1.5. Management of Protocol Imaging Performed Off-specification

Criteria should be included in the protocol that define acceptable, target, and ideal FDG-PET/CT imaging specifications and parameters. Imaging studies judged to be sub-optimal, if performed for "standard of care" could be repeated at the discretion of the site if the site deems the scan clinically unacceptable [4]. If the scan is judged unacceptable for research purposes, the study may be repeated as dictated by the protocol and informed consent. The protocol should then state how the cost of such repeated studies should be managed within the trial budget [6].

1.6. Management of Off-protocol Imaging

Acceptable: This UPICT protocol only addresses the performance of FDG-PET/CT in the context of a clinical trial. However, since imaging studies other than FDG-PET/CT might influence the conduct of the clinical trial including, but not limited to, the timing and performance of the FDG-PET/CT study(ies), the clinical trial protocol should explicitly state how all imaging tests, whether contemplated and/or obtained as part of the clinical trial or clinical care, should be managed with regard to the conduct of the trial. For the management of FDG-PET/CT studies performed off-schedule and/or outside of specifications, please see Sections 1.2 – 1.5.

1.7. Subject Selection Criteria Related to Imaging

Acceptable:
Fasting Blood Glucose: If quantitative FDG-PET/CT is to be used towards either primary, secondary, or exploratory aims, the study should include specific directions as to the management of subjects with abnormal fasting blood glucose measurements, whether known to be diabetic or not. While there is a paucity of scientific data to suggest the appropriate cutoff of blood glucose measurements that should be excluded from clinical trials that use FDG-PET/CT scan data, it is important to define how such subjects and the data from their imaging studies are managed to ensure comparability of imaging data within and among clinical trials. Specifically when quantitative FDG-PET/CT is being used as the study's primary endpoint, the acceptable

blood glucose range should be specified, as well as consideration and explanation as to the inclusion or exclusion of subjects with abnormal fasting blood glucose.

Lesion Conspicuity: It should be noted that tumors with low FDG uptake at baseline (also see Sections 9 and 10) may not be suitable for follow-up studies of treatment response with FDG-PET/CT (e.g. most FDG-avid tumor activity should be greater than 1.5 times hepatic mean +2 SD, see Section 10.2.1.1.2). Minimal lesion size and multiplicity may also be necessary as baseline inclusion criteria and if so those thresholds should be stated in the clinical trial protocol.

1.7.1. Relative Contraindications and Remediations

Inability to comply with or tolerate the performance of FDG-PET/CT imaging may be a relative exclusion criterion for subjects in a clinical trial that depends upon FDG-PET/CT for a primary or secondary endpoint. Examples of such relative contraindications include inability to remain motionless for the duration of the scan time or to lie flat for any number of reasons (e.g. severe congestive heart failure). However, such relative exclusion criteria are not unique to FDG-PET/CT. A plasma glucose level above the threshold as defined in Section 4.2.2 may necessitate the rescheduling of the FDG-PET/CT test to another day when the plasma glucose level is less than the defined threshold.

For this reason, subjects at risk for elevated plasma glucose levels should be scheduled early during the timing interval as specified in Section 1.2 so that if the test must be rescheduled the test date will still fall within the acceptable timing interval (See Section 1.2) so as to avoid a protocol deviation. In addition, it is suggested that for subjects who are known diabetics that three serial morning fasting blood glucose determinations (using home test kits) with values of less than 200 mg/dl (≈11.1 mmol/L) be obtained prior to scheduling the FDG-PET/CT test in order to assure that the test results may be valid within the context of the trial (see Sections 1.7.2, 3 and 4.2.2). Relative contraindications become absolute (i.e. Imaging Exclusion Criteria) when they cannot be remediated. When the FDG-PET/CT imaging endpoint is a trial endpoint, the subject would then be excluded from the trial.

1.7.2. Absolute Contraindications and Alternatives

The protocol should specifically define a threshold plasma glucose level that should represent an absolute exclusion criterion for participation in any clinical trial that depends on FDG-PET/CT imaging for any primary or a quantitative secondary endpoint if the plasma glucose level cannot be maintained below that threshold level using the diabetic management procedures as described in Section 4.2.2. Threshold plasma glucose levels for inclusion as suggested by referenced standards documents and publicly listed clinical trials include:

- A plasma glucose level: ≤126 mg/dl (≈7.0 mmol/L) [3]
- Blood glucose levels: ≤150 mg/dl (≈8.3 mmol/L) [6]
- Blood glucose levels: ≤200 mg/dl (≈11.1 mmol/L) [4, 11]
- Subjects known to be diabetic who have three serial fasting morning blood glucose levels of >200 mg/dl (despite adequate medical management) prior to the baseline or

initial FDG-PET/CT study should be excluded from a clinical trial in which quantitative FDG-PET/CT is used for a primary endpoint [14].

When FDG-PET/CT is used towards secondary and/or exploratory endpoints the trial should specifically state whether subjects with fasting blood glucose levels >200 mg/dl (≈11.1 mmol/L) will be included or excluded; and if included how the data from such subjects will be managed.

Furthermore, there are specific clinical trial purposes (e.g. progressive disease determination) for which fasting blood glucose levels >200 mg/dl (≈11.1 mmol/L) are acceptable.

Finally, there is a scientific gap in knowledge regarding the relationship between fasting blood glucose level and the effect on quantitative and qualitative FDG-PET/CT. It is recommended that investigators utilize pooled data from studies performed under rigorous protocols (such as the UPICT Oncologic FDG-PET/CT protocol) to investigate this relationship – including data from subjects with fasting blood glucose levels >200 mg/dl (≈11.1 mmol/L) [14].

Many clinical trials exclude subjects who are pregnant (or suspect they are pregnant) or breastfeeding when FDG-PET/CT is being used as a primary or secondary endpoint. However, such potential subjects may already be excluded on the basis of the index intervention under investigation without regard to the use of FDG-PET/CT.

Additional suggested exclusion criteria include weight exceeding table limits (300 - 450 lb or 136 – 205 kg for most current PET/CT scanners) and subjects with a history of life-threatening allergic / anaphylactoid reactions to any contrast media if contrast is being used in the study [4].

Relative contraindications become absolute (i.e. Imaging Exclusion Criteria) when they can no longer be remediated. When the FDG-PET/CT imaging endpoint is a trial endpoint, the subject would then be excluded from the trial.

1.7.3. Imaging-specific Inclusion Criteria

One source states that for clinical trials with longitudinal FDG-PET measurements as a primary endpoint might require a minimum tumor FDG-avidity based on the SUV (e.g. tumor SUV of > 1.5 x hepatic mean + 2 SD of hepatic mean using a 3 cm ROI to determine the mean) at baseline in order to remain on or to be eligible for participation on the study and have subsequent follow-up FDG-PET/CT scans [6, 15]. There may also be lesion "size" threshold (RECIST, WHO, volume) and/or lesion multiplicity (stage) threshold for eligibility (See also sections 9 and 10).

2. Site Selection, Qualification and Training (See also Section 12 relative to QC)

2.1. Personnel Qualifications

Acceptable: Each site shall have technical, physics, radiochemistry, and physician personnel trained in the use of FDG-PET/CT in the conduct of oncologic clinical trials prior to trial activation and subject accrual (or for **Target** Performance prior to site qualification).

In lieu of an on-site physicist, a consulting physicist or vendor-qualified service support personnel is acceptable.

2.1.1. Technical
Appropriate education, training, and certification of technologists is required to perform PET/CT. Representatives from the Society of Nuclear Medicine Technologist Section (SNMTS) and the American Society of Radiologic Technologists (ASRT) met in 2002 and published specific recommendations [16].

2.1.2. Physics
The SNM considers certification and continuing education in the appropriate sub- field(s) to demonstrate that an individual is competent to practice one or more of the subfield(s) of medical physics and to be a qualified medical physicist. The SNM recommends that the individual be certified in the appropriate subfield(s) by the American Board of Radiology (ABR) or the American Board of Science in Nuclear Medicine (ABSNM) [16].

2.1.3. Physician
Imaging experts interpreting PET/CT scans should have appropriate training in both PET and CT. A working group of representatives from the American College of Radiology, the Society of Nuclear Medicine (SNM), and the Society of Computed Body Tomography and Magnetic Resonance agree only appropriately trained, qualified physicians should interpret PET/CT [17]. This working group has also recommended the number of continuing medical education credits earned and the number of cases interpreted that would demonstrate adequate training [16].

2.1.4. Other (e.g., radiochemistry, radiobiologist, pharmacist, etc.)
Acceptable: For oncologic FDG-PET/CT the qualifications of the personnel involved in the preparation of the FDG should be appropriate to comply with the FDA part 212 specifications or the international equivalent, as appropriate to the regulatory jurisdiction within which the FDG will be administered.

2.2. Imaging Equipment

At this time, the current protocol is exclusive for PET/CT scanners used in whole-body oncology trials using. While PET/MR is an emerging technology, further evaluation and validation is necessary and should be addressed separately in the context of the special clinical trial. Each site needs to have contemporary PET/CT system(s) [14]. Multiple references suggest that integrated PET/CT scanners are preferable to be used for imaging based on increased accuracy for lesion localization and characterization than that obtained from the results obtained from PET and CT separately and interpreted sided by side or following software based fusion of the PET and CT datasets [3]. PET scanners that utilize NaI detectors are excluded [2, 8].

An important aspect of quantitative multi-center PET imaging studies and therefore integral to the qualification of imaging platforms is the cross-calibration of scanner performance across various imaging sites. Several societies, organizations and clinical trials networks, such as the NCI, ACRIN, EORTC, EANM and SNMMI, etc., have developed multi-center clinical trials imaging guidelines and have set up or are setting up PET/CT system validation and site accreditation

programs to ensure that data collected using these systems are comparable, i.e. can be exchanged.

These site accreditation programs use different phantoms for this purpose; among the performance characteristics that are tested are:

1) the verification of a correct (cross-) calibration of the PET/CT system (against a dose calibrator) [3, 8, 9, 11].
2) scanner normalization and uniformity [8].
3) the assessment of 2D or 3D SUV recovery coefficients (thereby essentially assessing contrast recovery and/or partial volume effects as a function of sphere size or rod diameter) [3, 9, 11].

Despite the differences in the implementation of scanner validations, all site accreditation programs aim to assess image quality on some or all of these main image characteristics. Future work should focus on further aligning of the activities of these societies, either by harmonizing the scanner validation platforms/phantoms and development of an equivalent scanner multi-center QC program. The latter should be feasible considering the good agreement between the societies regarding the image characteristics to be verified. At present, there is a strong interest from all groups in establishing a common FDG PET standard.

Site qualification by a standardized method (including, but not limited to, documentation of a rigorous quality control program incorporating the use of a uniform phantom to verify scanner normalization and calibration) is the minimum acceptable for clinical trials [8] and use of a standardized multi-compartmental phantom (to additionally evaluate detectability, resolution and contrast recovery) at all sites for this purpose is the target [14]. For a detailed discussion with materials and methods, see Section 12.1.

Initial and ongoing periodic QC for CT as used for attenuation correction and localization is included within the scope of this document (see 12.1 for detail). However, QC for diagnostic CT performed in conjunction with oncologic FDG-PET/CT is not included within the scope of this document. Documentation for diagnostic CT may be obtained from other UPICT documents.

The sites also need to have all the ancillary equipment for conduct of the trial including, but not limited to, appropriately calibrated glucose measuring device, dose calibrators, stadiometer to measure height, and scales to weigh subjects. See Section 12.1.1 for quality control.

2.3. Infrastructure

Acceptable: All sites participating in the conduct of an oncologic clinical trial utilizing FDG-PET/CT must have oversight by an Institutional Review Board, Ethics Committee, or equivalent group that oversees and is permitted to review and approve experimental studies involving human subjects; a Radiation Safety Committee or equivalent body; and an entity designated to oversee the privacy of personal healthcare information (e.g. HIPAA Board or equivalent; n.b. in many United States institutions the IRB serves as the Privacy Board for research matters). The participating site must also have the prerequisite infrastructure to perform the specified acquisition, archival, de-identification, and transfer of imaging data as required by the clinical trial protocol in a matter compliant with the protocol and all local, regional, and national regulatory

requirements. Sufficient infrastructure must be demonstrated and documented to perform and report the quality control procedures specified within the clinical trial protocol with expectations enumerated in the clinical trial within the appropriate documentation.

2.4 Quality Control {*This section intentionally omitted*.}

2.5. Protocol-specific Training {*This section intentionally omitted*.}

3. Subject Scheduling

Prior to scheduling potential and/or already accrued subjects for FDG-PET/CT with its inherent (albeit minimal) risks, confirmation of appropriateness for imaging (e.g. history, physical examination, staging, biopsy for diagnosis, etc.) should be performed and documented. Scheduling diabetic subjects may require special attention (please see Section 4.2.2 for additional details) and therefore this should be specifically queried at the time of scheduling. At the time of scheduling, the study team should determine that inclusion of the subject does not violate any of the study-specific inclusion and exclusion criteria pertinent to the FDG-PET/CT study. (SNM GHS) For considerations related to the scheduling of subjects who are known to be diabetic, please also see Sections 1.7.2 and 4.2.2.

- Additional scheduling recommendations for diabetic subjects are suggested by two references [3, 11]. These include the following:
    - For type I diabetes:
        - Ideal to achieve euglycemia prior to PET study
        - Schedule study for late morning by eating normal breakfast at 7 am and taking normal amount of insulin; then fast for at least 4 hours till exam
    - For type II diabetes:
        - Schedule study for late morning
        - Comply with at least 4 hour fast till exam
        - Continue oral medication (hypoglycemic) as usual
- One reference suggests the following for diabetic management:
    - Diabetic subjects should be scanned early in the morning before the first meal, and doses of insulin and/or hypoglycemic medication should be titrated appropriately in consultation with the subject's referring physician [18].

Before scheduling an FDG-PET study, diabetic subjects should test their ability to maintain reasonable plasma glucose levels after fasting, while avoiding insulin close to the time that FDG would be administered.

- For known diabetic subjects with anticipated fasting blood glucose (FBG) measurements for the day of the examination between 126 mg/dl (≈7.0 mmol/L) and 200mg/dl (≈11.1 mmol/L), the following scheduling recommendations apply:
- Ideal / Target: Type I and Type II diabetic subjects should be scanned early in the morning before the first meal, and doses of insulin and/or hypoglycemic medication should be withheld if glucose levels remain in the acceptable range. This should be established from morning blood glucose levels prior to the study.
- Acceptable: Type I and Type II diabetic subjects, who cannot reliably attain acceptable glucose levels early in the morning, should be scheduled for late morning, and should eat a normal breakfast at 7 am and take their normal morning diabetic drugs; then fast for at least 4 hours till exam. This strategy is acceptable only for:

  - Non-quantitative PET/CT, or
  - Endpoints that are not for the primary aim, or
  - Subjects whose baseline study was performed with a FBG <200 mg/dl (≈11.1 mmol/L), but who have become uncontrolled hyperglycemics secondary to treatment effect, disease progression, or are being evaluated for exploratory endpoints
- In each case, the goal is to achieve a fasting blood glucose with the prescribed range (e.g., ≤126 (≈7.0 mmol/L), ≤150 (≈8.3 mmol/L), or ≤200 mg/dl (≈11.1 mmol/L) dependent on the clinical status of the subject, mechanism of therapy, and the utility of the FDG-PET/CT test in the clinical trial) [14]

3.1. Timing Relative to Index Intervention Activity
Acceptable: Please see Section 1.2.

3.2. Timing Relative to confounding Activities *(to minimize "impact")*

Activities, tests, and interventions that might increase the chances for false positive and/or false negative FDG-PET/CT studies should be avoided prior to scans. The allowable interval between the potentially confounding event and the imaging test will be dependent on the nature of the confounder. For example, a percutaneous or excisional biopsy of a suspicious mass may cause focally increased FDG-PET activity or might lead to the appearance of a non-malignant mass (e.g., hematoma) on the CT portion of the study. A percutaneous ablation procedure of a known malignant focus may cause focally increased FDG-PET activity and/or an immediate post-ablation increase in the apparent volume of the ablation target lesion. The time of onset and the duration of the increased FDG-PET activity and/or the change in lesion volume might be different for these two different confounding factors.

If iodinated contrast is to be used for the CT portion of the PET/CT study, conflict with other tests and treatments should be avoided congruous with community standards of care (e.g. thyroid scan).

3.3. Scheduling Ancillary Testing

Avoid scheduling tests that might confound the qualitative or quantitative results of the FDG-PET/CT study within the time period prior to the scan. For example, a glucose tolerance test should not be scheduled during the 24 hours prior to the performance of FDG-PET/CT. Similarly, other tests that might involve increasing plasma glucose, insulin, or corticosteroid levels should also be avoided. Exercise cardiac stress testing should be avoided during the twenty-four (24) hours prior to the performance of FDG-PET/CT. Similarly, other tests that might involve vigorous exercise and thereby increase muscle metabolic function should also be avoided.

4. Subject Preparation

4.1. Prior to Arrival

The main purpose of subject preparation is to reduce background tracer uptake in normal tissue (kidneys, bladder, skeletal muscle, myocardium, brown fat) while maintaining and optimizing tracer uptake in the target structures (tumor tissue) [19]. Below is a generally applicable protocol to address (1) Dietary, (2) Fluid Intake, and (3) Other activities that may impact the FDG-PET/CT procedure or results.

(1) Dietary (for the management of previously known or unknown diabetic subjects, please see section 4.2.2):

- According to two sources, subjects should fast for an absolute minimum (acceptable level) of 4 hours prior to FDG injection [18], although the target pre-test fasting period is recommend as a 6 hour minimum [3, 11]. This can be achieved as follows:
    - Subjects scheduled to undergo the PET study in the morning should not eat after midnight and preferably have a light meal during the evening prior to the PET study.
    - Subjects scheduled for an afternoon PET study may have a light breakfast before 8 am.
    - Medication can be taken as prescribed (see Section 4.2.2 for diabetic management)
- Two sources have stated that a low carbohydrate diet should be followed for 24 hours before the study, culminating with fasting for the final six hours [2, 18].
- Enteral nutrition is at least six (6) hours prior to the anticipated time of FDG administration [14].
- One study has suggested that a high-fat, low-carbohydrate meal is preferred for the last meal prior to commencing the period of fasting [20, 21]; Although there are insufficient data to recommend these strategies as routine at this time [14].

(2) Fluid Intake:
Adequate hydration (before and after FDG administration) is important (both to ensure a sufficiently low FDG concentration in urine (less artifacts) and for radiation safety reasons). Whichever hydration strategy is used (how much and when to administer), the protocol should be uniform among sites during a trial.

Specific hydration recommendations include:

- oral intake of at least 710-1665 ml of water while fasting [2]
- consumption of two to three 8-12 oz water (710-1065 ml) while fasting [18]
- one liter during 2 hours prior to FDG administration [3, 11]

If IV contrast is to be injected as part of the study, subjects should be asked to drink more fluid (total of 1 liter) during the two hours prior to the study. The fluid administered should not contain glucose or caffeine. It is acceptable for subjects to receive non-glucose containing IV solutions such as normal or dilute saline. Lactated Ringer's solution is not acceptable and should be discontinued. This hydration strategy should be modulated as clinically appropriate in subjects with certain medical conditions including, but not limited to congestive heart failure, renal failure and fluid retention for example [14].

Parenteral nutrition and intravenous fluids containing glucose should be discontinued at least 4 (acceptable) - 6 (target) hours before the PET examination [3, 14, 18]. The infusion used to administer intravenous pre-procedural hydration must not contain any glucose.

(3) Other Activities:

To minimize uptake of radiotracer into muscle, the subject should avoid strenuous exercise, or cold exposure before the PET exam for a minimum acceptable period of at least 6 hours but preferably for a target time period of 24 hours prior to the PET exam [2, 3, 18].

Other activities that might be avoided are contained in sections 3 and 3.2.

Performing FDG-PET scanning in the context of recent (within 24 hour) steroid administration may affect the subject's glucose control and hence SUV quantitation. Consequently, if intravenous contrast enhanced CT is required by the protocol in addition to the PET/CT exam, then special consideration is needed for subjects with iodinated contrast allergy who will require steroid premedication for the contrast enhanced CT. In this situation it is preferable that the contrast enhanced CT scan (with appropriate steroid administration) is performed at least one to two days following the 'non-contrast' PET/CT exam. If steroid premedication is given prior to PET/CT exam, then the quantitative assessment obtained from the PET exam may be adversely affected. In cases where premedication is needed for the contrast enhanced CT, the local imaging facility's premedication strategy should be followed and used consistently for the subject across all time points.

4.2. Upon Arrival

4.2.1. Confirmation of subject compliance with instructions

Upon arrival 1) confirmation of subject compliance with pre-procedure instructions and 2) the occurrence of potentially confounding events should be documented on the appropriate case report forms.

The documentation should include some or all of the following:

- timing, character, and amount of the most recent previous oral and/or intravenous intake of fluid and nutrients
- timing and dosages of relevant non-prescription and prescription medications taken prior to the PET/CT scan (e.g., the last cycle of chemotherapy or non-cytotoxic pharmacotherapy, administration of growth factors, cytokines, steroids, beta blockers, etc.)
- extent of physical activity and most recent exposure to cold temperature for the preceding 24 hours
- timing and description of medical procedures performed prior to the PET/CT scan (e.g., radiation therapy, biopsy, surgery)
- timing and description of relevant medical tests performed prior to the PET/CT scan (e.g., invasive tests and/or tests that involve the administration of exogenous substances and/or tests that involve vigorous physical activities)
- timing of iodinated contrast reaction prophylaxis if appropriate
- Confirmation that the subject has completed the trial Informed Consent Document.

The FDG-PET/CT procedure should be explained to the subject and exam-specific consent should be obtained if that is the standard of care for the site or the standard established for the specific clinical trial. There should be documentation of subject-specific risk factors

including, but not limited to, previous contrast reactions (if iodinated contrast is to be used) and the presence of implanted electronic devices (e.g. pacemakers, neural stimulators, cochlear implants) [18].

4.2.2. Ancillary Testing To Be Performed Upon Arrival

Subject height and body weight must be measured precisely with standardized measurement devices and with the subject in a gown or light clothing and recorded as the minimum acceptable standard [2, 3, 11, 18]. The target standard would add that for serial studies in the same subject, weight should be measured directly prior to each PET study since body weight often changes during the course of the study [3, 14].

Blood glucose monitoring, measurement and documentation and the appropriate management/disposition of hyperglycemic/ diabetic subjects are addressed by all references and should be included as a minimum acceptable standard of performance.

- It is important to measure and document subject blood glucose level shortly prior to and target within the 2 hours prior to (ideally within 1 hour for all subjects and target within 1 hour for insulin-requiring diabetic subjects) FDG administration (all, SNM GHS).
- Ideal: fasting blood glucose level < 126 mg/dL (≈7.0 mmol/L).in the absence of recent insulin therapy. This may have the effect of excluding diabetic subjects, including those who are undiagnosed at the time of the scan.
- Target: fasting blood glucose level < 150 mg/dL (≈8.3 mmol/L).
- Acceptable: Subjects with blood glucose measurements between 126 mg/dL (≈7.0 mmol/L) and 200mg/dL (≈11.1 mmol/L) can be imaged [2, 3, 11, 18], there are varying actions suggested by the different references.
  - There is no consensus from these references for diabetic or non-diabetic subject management in the glucose range of 126 - 200 mg/dL (≈7.0 - 11.1 mmol/L).
  - The imaging protocol for each individual clinical trial should indicate the glucose cut-off thresholds and the exact management for diabetic and non-diabetic subjects with plasma glucose levels between 126 - 200 mg/dl (≈7.0 - 11.1 mmol/L), especially if the quantitative data from the FDG-PET/CT examination will be used towards a primary or secondary endpoint and/or will be compared in a serial manner over the course of the protocol.
  - Subjects with blood glucose level > 200 mg/dL (≈11.1 mmol/L) should be rescheduled. Adjustments to diet, medications, and exercise made if necessary, so that the fasting blood glucose concentration can be brought down to the acceptable range at the time of FDG injection, or excluded depending on the subject circumstances and the trial being conducted [3, 5, 18].
  - Consideration should be given to handling the data differently from subjects if there is significant (i.e > 50%) change in fasting blood glucose between baseline and follow up studies.

- Secondary to recognized problems with administration of insulin (due to alteration of FDG biodistribution and diminished accuracy of SUV determination-NCI), insulin must not be given to reduce pre-FDG-administration glucose levels, unless the interval between administration of insulin and FDG is more than 4 hours [2, 3].

Commented [MMG1]: This sentence was added after discussion within the writing group. It was a concept that had been present in earlier versions of the document but had been deleted.

4.2.3. Preparation for Exam

In order to avoid artifactual distribution of the FDG, it is critical that subject preparation, after arrival and prior to imaging, are standardized among all sites and subjects throughout the conduct of the clinical trial [2, 3, 11, 12, 18].

- The waiting and preparation rooms should be relaxing and warm (> 75° F or 22° C) during the entire uptake period (and for as long as reasonably practicable prior to injection, at least 15 minutes is suggested as acceptable. Target time is 30 minutes). Blankets should be provided if necessary [14].
- In addition to a warm room, several studies have shown that one option to reduce brown fat uptake is beta blockade such as the administration of propranolol [22, 23]. More recent studies have shown that for patients 21 and under, a lower dose of 0.33 mg/kg with a maximum of 20 mg administered one hour before FDG injection has been effective. For adult patients with a history of brown fat uptake, 20 mg has also been used.
- The subject should remain recumbent or may be comfortably seated; activity and conversation should be kept to an absolute minimum. For example, the subject should be asked to refrain from speaking, chewing, or reading during the uptake period [14]. For brain imaging the subject should be in a room that is dimly lit and quiet for FDG administration and subsequent uptake period [18].
- The subject may use the rest room and should void immediately (5 – 10 minutes) prior to the FDG-PET/CT image acquisition phase of the examination.
- Bladder catheterization is not routinely necessary; but if necessary the catheter should be placed prior to injection of FDG. Bladder catheterization may be important for the evaluation of pelvic tumors (e.g. cervix or prostate cancer).
- Following the administration of FDG, the subject should drink 500 ml (or 8 – 12 oz, 237-354 ml per ACRIN [18]) of water (or receive by intravenous administration 250 - 500 ml of non-glucose containing fluid). Fluid intake may need to be modified for those subjects on fluid restriction.
- For specific areas of anatomic interest (e.g. tumors located in the lower abdomen, pelvis or kidney) intravenous diuretic agents may be used (e.g., 20 – 40 mg of furosemide given nearly contemporaneously (within 10 – 15 minutes) with the administration of FDG). Per the SNM harmonization summit [14], if bladder catheterization is performed IV diuretics should be administered as described herein so as to ensure that the concentration of activity in the renal collecting systems and bladder is relatively dilute.
- Sedation is not routinely required, but is not contraindicated provided that the sedative used does not interfere with the uptake of FDG. If sedation might be used, the subject should be instructed in advance that operation of a motorized vehicle will be prohibited after the FDG-PET/CT test. Sedation may have utility in specific clinical circumstances such as brain or head and neck tumors, claustrophobic subjects, or children.
- The amount of fluid intake and use of all medications (e.g. diuretic, sedative) must be documented on the appropriate case report form.
- Subjects undergoing a CT scan should empty their pockets and remove any clothing containing metal and any metallic jewelry from the body parts to be scanned, changing into a hospital gown if necessary.[17]

5. Imaging-related Substance Preparation and Administration

IV and oral iodinated contrast is not discussed as part of this document as its utility is related to the diagnostic CT examination.

FDG should be of high quality and purity. For example, the FDG radiopharmaceutical must be produced under Current Good Manufacturing Practice as specified by the FDA, EU, European Pharmacopeia or other appropriate national regulatory agency. U.S. regulations such as 21CFR212 or USP<823> Radiopharmaceuticals for Positron Emission Tomography must be followed in the U.S. or for trials submitted to US Regulatory. For example, in the US, for clinical practice, FDG production under NDA or ANDA or under IND for research purposes is mandatory. The quality control should be consistent with Section 12.4. If IV and/or oral iodinated contrast is to be used in the study, the density, quantity, and composition (if pertinent) should be specified in the protocol.

5.1. Substance Description and Purpose

A brief statement regarding FDG as the imaging agent should be included in the clinical trial protocol where appropriate; for example: FDG is a glucose analogue. Its use in oncology is based on the fact that most types of tumors utilize more glucose than most other types of normal tissue.

5.2. Dose Calculation and/or Schedule

The FDG dose is usually around 5mCi in Europe [3] and between 10mCi (=370 MBq) [12] and 20 mCi (=740 MBq) [18] in the United States. Further FDG dose refinement and/or dose reduction can be achieved by taking into account: (1) patient weight, for example by applying a dose of 5 – 8 MBq/kg; (2) 2D versus 3D scanning mode; (3) acquisition time per bed position and; (4) percentage bed overlap of subsequent bed positions. The exact dose and the time at which dose is calibrated should be recorded. Residual dose remaining in the tubing, syringe or automated administration system and any dose spilled during injection should be recorded.[3, 11, 12, 18]

- In the case of using an automated system, the administered FDG activity should be within 3% accuracy (this must be ensured by manufacturer and verified by the user); i.e., the actual administered activity may not deviate more than 3% from that indicated by the reading of that device or dose calibrator following instructions given by the manufacturer of the automated administration system.
- Residual activity as determined by the above methods should be used to correct the administered dose for any quantitative results reported.

Any upper dose limits related to dead time/count rate limitations, as recommended by the tomograph manufacturer should be taken into account. Moreover, (upper) dose limits may apply because of national or local legislation. In case upper dose limits apply, consistent image quality across sites should be accomplished by increasing scanning time. For pediatric studies, other guidelines may apply, such as the EANM pediatric dose card [24, 25].

5.3. Timing, Subject Activity Level, and Factors Relevant to Initiation of Image Data Acquisition

FDG uptake into both tumors and other body tissues is a dynamic process that peaks and plateaus at various time points dependent upon multiple variables [26, 27]. Therefore, it is extremely important that (1) the time interval between FDG administration and the start of emission scan acquisition is consistent and (2) when repeating a scan on the same subject, it is essential to use the same interval after injection for scans performed at different times.

The suggested consensus time (from all references) between FDG administration and scan acquisition is 60 minutes based on historical use of this test; assuming this is the target window, an acceptable window is often cited as +/- 5 minutes (55-65 minutes). Two references allow the acceptable window to be +/- 10 minutes (50-70 minutes), which is considered the absolute minimum of acceptability [2, 18, 28].

However, on the basis of the SNM harmonization summit while the "target" tracer uptake time should be 60 minutes, there was consensus that the "acceptable" window should be from 55 to 75 minutes so as to ensure that imaging does not begin prematurely so as to allow adequate tumor uptake of FDG and to account for the practicality of work flow which often does not accommodate imaging at exactly 60 minutes after FDG injection [14]. The exact time of injection must be recorded; the time of injection initiation should be used as the time to be recorded. Ideally, the injection and flush should be completed within one minute with the rate of injection appropriate to the quality of the vein accessed for FDG administration so as to avoid compromising the integrity of the injection vein.

More recent evidence might justify a target interval of greater than 60 minutes for a particular trial. If a target time greater than 60 minutes is chosen for a specific trial, the imaging protocol should justify the specific time chosen, as well as the acceptable window about this target time. Furthermore, as routine clinical practice might not allow the use of pre-recruitment scan for the study, the protocol should include a plan for repeating the baseline scan if necessary to allow appropriate inter-time-point comparisons [6, 14].

When repeating a scan on the same subject, especially in the context of therapy response assessment, it is essential to apply the same time interval with target window of +/- 10 minutes (with an acceptable window of +/- 15 minutes) provided that the scan must not begin prior to 55 minutes after the injection of FDG [14]. If a limited or targeted scan is obtained at follow-up after a whole body scan was performed at baseline, one should consider adjusting the timing of the follow up scan to be congruent with the timing for the same anatomic region as achieved during the baseline study.

If, for scientific reasons, an alternate time (between dose administration and scan acquisition) is targeted for a specific protocol, then the rationale for this deviation should be stated; inter-time point consistency must still be followed [2].

5.4. Administration Route

FDG should be administered intravenously through a large bore (≥21 gauge) indwelling catheter placed anatomically remote (e.g. contralateral extremity to site of disease if at all possible) to any site(s) of suspected pathology, preferably in an antecubital vein. Intravenous ports should not be

used, unless no other venous access is available. If a port is used additional flush volume should be used. As reproducible and correct administration of FDG is required for quantitation purposes, extravasation or peri-venous administration should be avoided [2, 3, 11, 18]. If an infiltration is suspected, the fact should be recorded and if the study is quantitative, i.e. SUVs will be measured, then the infiltration site should be imaged and the approximate amount of infiltration should be calculated. If the infiltration is greater than 5% of the administered dose and the quantitative result from the FDG-PET/CT study is a primary or secondary endpoint, the data point might be censored from review or the subject might not be included in study [14]. The injection site should be documented on the appropriate case report form [18].

Presuming that the IV access site is properly functioning, the same route of administration may be used for iodinated contrast as is used for FDG.

5.5. Rate, Delay and Related Parameters / Apparatus

Either manual or automated injection systems may be used to administer the FDG.

- In the case of manual administration, a three-way valve system should be attached to the previously placed intravenous cannula (See Section 5.4) so as to allow at least a 10 cc normal (0.9% NaCl) saline flush following FDG injection. Residual activity within the syringe, and as much of the administration system as is available (including the needle cap) must be measured and the residual dose should be documented (See Section 5.2) [3, 18, 28].
- In the case of an automated administration system, the manufacturer's instructions should be followed. However, the automated system and administration procedures must be ensured by the manufacturer and verified by the user to perform within the characteristics specified in Section 5.2).

5.6. Required Visualization / Monitoring, if any – **NA**

5.7. Quality Control -- See 12.21.

6. Individual Subject Imaging-related Quality Control -- See 12.3.

7. Imaging Procedure

7.1. Required Characteristics of Resulting Data

7.1.1. Data Content

For most Oncology indications, anatomic coverage should include from the skull base (external auditory meatus to the proximal to mid-thigh. This is considered a 'whole body' scan. However, other ranges could be used as appropriate for specific clinical trials. However, the clinical trial should then provide specific instructions with justification. Usually the scanning direction should be caudiocranial to minimize effect from increasing bladder activity during the scan. Scanning direction should be protocol specified. It is critical that for a given subject, scanning direction on baseline scans be duplicated at follow-up time points [2, 14].

Any potential sources of artifact (e.g. urine collection bags, surgical drainage bags, IV lines and related devices) should be managed or positioned so as to eliminate or minimize degradation of the image and image-related data.

Extended anatomic coverage (e.g. brain or extremities) may be performed for tumors that show higher probability of metastasis or direct extension above the skull base or below the mid-thigh. If extended anatomic coverage is performed, this could be performed as a continuation of the skull base to mid-thigh exam or be performed as a two-step protocol. Two-step exam may be preferable, especially in the case of head and neck tumors. If a two-step or an anatomy extended examination is performed, attention to scan timing is critical to provide time relevant comparison with earlier time points (see section 5.3).

Either one of the following two different scanning strategies can be used for FDG-PET/CT acquisition. For the first strategy, there is no intent to obtain a diagnostic CT scan at the FDG-PET imaging session; for the second strategy, a diagnostic CT is obtained. Whichever strategy is used, it is recommended that all FDG-PET/CT scans for an individual subject (**target for all subjects**) be performed using the same strategy for all sequential time points. The workflow chosen should be described in the protocol and should be tailored commensurate to the level of expectation of the obtained data (e.g. qualitative or quantitative SUV analysis).

Strategy 1: For FDG-PET/CT in which the CT is used for attenuation correction and localization only (no diagnostic CT intent):

- CT Scout (topogram), followed by
- CT for anatomic localization and attenuation correction, followed by
- Emission scan acquisition

Strategy 2: For FDG-PET/CT in which a diagnostic CT is performed in conjunction with FDG-PET, one of two strategies shall be used. Either (2a) follow Strategy 1 and then, with no or minimal patient motion after the PET Emission scan acquisition, perform an additional IV contrast-enhanced diagnostic CT or (2b) perform a contrast-enhanced diagnostic CT before following the workflow described in Strategy 1.

For both strategies, there are several common issues specific to the CT exam that may have an impact on quantitative FDG-PET output, which need attention and protocol specification. These include: (1) contrast material administration, (2) respiratory motion compensation instructions and (3) CT scanning technique (kVp, mAs and pitch).

All these issues should be addressed in the clinical trial protocol, (with target of consistency across all time points for each given subject and ideally with consistency across all sites and all subjects (both inter-subject, and intra- and inter-facility). The actual details of imaging for each subject at each time point should always be recorded.

Any particular clinical trial should NOT allow some sites to implement one strategy and other sites to implement the alternative.

For strategy 1 where the CT is used for attenuation correction and localization only (no diagnostic CT intent), the following behavior levels apply:

- Contrast Material

  The presence of a positive contrast agent (IV or oral), by affecting the CT attenuation map, can result in a small variability of quantitative SUV evaluation. If this were the only consideration, then the ideal would be to prohibit CT contrast administration. However, in some clinical situations (dependent upon tumor type, tumor behavior or level of anatomic interest), the benefit of oral CT contrast may outweigh the small errors induced in SUV measurement that may include increased SUV variability. Consequently, ideal and target approaches are grouped as below. Each protocol should specify the desired approach for the given study. Most importantly, for each subject, the same approach should be followed for all imaging time points.

  a. Acceptable
  No IV contrast; dilute positive oral contrast is acceptable

  b. Target/Ideal
  No positive contrast agent (IV or oral) for FDG-PET/CT studies with a predominant intent of quantitation at both baseline and follow-up

  No IV contrast agent; negative or dilute positive oral contrast is allowed for FDG-PET/CT studies with primary quantitative intent with additional need for oral contrast to increase confidence of true positive disease detection and/or additional qualitative assessment.

- Respiratory Motion Compensation

  Respiratory motion causes SUV errors by two mechanisms: motion blurring and attenuation correction mismatches between CT transmission map and emission data.

  a. Acceptable
  Verbal instruction to the subject for shallow breathing during CT and PET.

  b. Target
  Verbal instructions to subject for similar shallow breathing during both the PET and CT acquisitions; respiratory gating if called for in a given protocol specification

  c. Ideal
  Verbal instructions to subject for similar shallow breathing during both the PET and CT acquisitions; respiratory gating if called for given protocol specification; possibly with advanced methodologies for respiratory synchronization if offered by manufacturer and appropriate to the study. Respiratory gating on PET may require several CT attenuation maps for optimal quantitation.

- CT Technique

  a. Acceptable
  Recording of actual kVp and exposure (CTDI, DLP) for each subject at each time point. CT dose exposure should be appropriately reduced in smaller patients and children.

b. Target

Consistency in use of kVp and low exposure (CTDI, DLP) for all time points for a given subject in addition to the Acceptable conditions stated below. CT dose exposure should be appropriately reduced in smaller patients and children.

c. Ideal

Use of manufacturer recommended kVp and exposure CT Dose Index (CTDI) or Dose Length Product (DLP) settings for low dose exam in addition to the Target and Acceptable conditions stated below. CT dose exposure should be appropriately reduced in smaller patients and children.

Regarding CT radiation exposure, rules of "As Low as Reasonably Achievable" (ALARA) should be followed. For a given protocol, the purpose of performing the CT scan (attenuation correction only or attenuation correction and anatomic localization) should be determined.

The CT technique (mAs, pitch, collimation, kVp, and slice thickness) used should result in as low as reasonably achievable exposure needed to achieve the intended goal of imaging working with the scanner manufacturer to achieve this objective. The technique used for an imaging session should be repeated for that subject for all subsequent time points assuming it was properly performed on the first study.

Strategy 2: For FDG-PET/CT in which a diagnostic CT is performed in conjunction with FDG-PET, since there may be variability introduced into the SUV calculations by the presence of even dilute intravascular iodinated contrast. Consequently, each clinical trial should choose either the Acceptable or the Target/Ideal strategy as described below for use at all sites, for all time points, and for all subjects. Any particular clinical trials should NOT allow some sites to implement one strategy and other sites to implement the alternative.

a. Acceptable

Perform a contrast enhanced (IV and dilute or negative oral contrast) diagnostic CT before step 1 of Strategy 1, then with no or minimal patient motion between the diagnostic CT and the PET/CT complete steps 1-3 (including a separate tidal-breathing AC / localization CT) of Strategy 1 ensuring that the diagnostic CT acquisition is performed consistently for a given subject across all time points. The IV contrast would then be in equilibrium phase during the emission scan acquisition and the AC / localization CT scan. (note – since there are no data as to the magnitude of variance in SUV calculation between the Ideal / Target strategy and the Acceptable strategy, perhaps QIBA should investigate if the Acceptable strategy is indeed truly acceptable for quantitative FDG-PET/CT in the conduct of a clinical trial.)

b. Target / Ideal

Follow Strategy 1 (steps 1-3 above) and then with no or minimal patient motion between the diagnostic CT and the PET/CT perform an additional IV contrast-enhanced diagnostic CT after the emission PET scan acquisition. Ensure that the

diagnostic CT acquisition is performed consistently for a given subject across all time points. Note that for this case, use negative or dilute positive oral contrast for the non-attenuation CT scan.

In some instances, such as head and neck cancer, a separate dedicated PET and CT acquisition may be appropriate with the arms in a different position (down) than would be used for the remainder of the whole body study (see also Section 7.2.1 "Subject Positioning").

c. Unacceptable

Performance of a single diagnostic quality CT study prior to or after the emission scan for all purposes (i.e., anatomic localization, attenuation correction, and diagnostic CT information) is considered unacceptable for clinical trial use. The major negatives for this strategy are due to mis-registration and incorrect attenuation correction application (especially around the level of the diaphragm) due to differential diaphragmatic position between optimal diagnostic CT (typically full breath hold inspiration) and emission (tidal breathing) FDG-PET scan acquisitions. [29]

This is believed to strongly outweigh the benefit of radiation dose reduction achieved by eliminating the low-dose CT for anatomic localization / attenuation correction map. A dose reduction can be achieved in cases in which a diagnostic IV contrast CT is required, by limiting the CT with contrast to the most relevant regions of the body, which may be a smaller extent than the area imaged on PET.

7.1.2. Data Structure

Acceptable / Target: The matrix size, slice thickness, and reconstruction zoom should yield a target reconstructed voxel size of 3 – 4 mm in all three dimensions (i.e., not achieved through post-processing), although not necessarily isotropic. For QC, see section 12.1.1.

Ideal: Reconstructed voxel size (i.e., not achieved through post-processing) should be as small as possible without introducing artifacts and also so as to be consistent across all trial sites; with current technology 2 – 3 mm in all three dimensions is achievable.

7.1.3. Data Quality

Image quality (as defined by SUV calibration, SUV Recovery Coefficient, and SNR) should be such that when applying the same acquisition and reconstruction protocol as used in subject scanning to the protocol specified phantom(s) the output should meet the QC standards as stated in Section 12.1.1.

Treatment response assessment and classification (based on criteria) require several quantitative and qualitative assessments. For details see Sections 9 and 10. In summary, however, the analysis and interpretation steps depend on several aspects including, but not limited to, assessment of lesion eligibility, percentage change in activity of specified lesions

at each time point relative to baseline, and the appearance of new lesions that meet eligibility criteria.

For the first two aspects (lesion eligibility and measuring percentage change) standardization of quantitative image quality, e.g. by means of harmonizing recovery coefficients measured in specific dedicated phantoms, will result in more uniform lesion selection and response assessments across institutes. Consequently, harmonizing quantitative performance of PET/CT systems coupled with defining some minimum and/or optimum performance metrics should be a strong consideration in the design of a multicenter trial.

For the assessment of progression related to the appearance of one or more new lesion(s), it is important to set a minimal threshold for image quality with respect to lesion detectability. As such, scanners need to have a minimal image quality performance/lesion detectability/SNR in order to be suitable to be used in trials. It therefore is conceivable that two different sets of reconstruction algorithms and settings may be necessary to use in the trial; one for lesion detection and the other for lesion quantitation.

Both lesion detectability and quantitation must be carefully considered during study design so as to properly define minimum quality standards to be applied across all sites and scanner platforms (see Section 12.1.1).

7.2. Imaging Data Acquisition

All QC procedures should be followed and documented prior to the initiation of acquisition.

For serial scans of the same subject, every attempt should be made to use the same scanner, and the same scanner model throughout the trial [28].

However, in some cases a different scanner that has been previously qualified and is the same platform as the scanner used at baseline can be used for a subject's follow-up scan in the instance of equipment malfunction [6].

The ideal level of performance is that all serial scans on a subject should be performed on the same scanner with the same software version; acceptable / target performance is that all serial scans on a subject should be performed on equivalent scanners (i.e. the same model) but also with the same software version) [14].

Additionally, all scan acquisitions for a given subject should include identical transmission and emission scanning techniques and emission scan duration per bed position [28]. There is no consensus provided on emission scan time range. The number of bed positions and the acquisition time per bed position will be scanner specific. Typical parameters are 6 bed positions and an acquisition of 2 – 5 min per bed position.

The minimum acceptable time per bed position should be between 2 and 4 minutes for a 3D acquisition with 2D acquisitions typically requiring at least 1.5 - 2x longer depending on the administered FDG dose; although the absolute impact on image quality by scan time per bed position is currently undefined it is dependent on several pertinent factors including, but not

limited to, administered dose, body weight and habitus, bed overlap, and specific model / version of the imaging platform used. In general, increased scan time per bed position will improve the SNR and thus it may be important to increase scan time when quantitative metrics are used towards a primary endpoint.

One vendor has implemented continuous bed motion acquisition to provide greater flexibility in defining anatomic scanning range and to improve the uniformity of axial noise variance. Bed velocity should be adjusted according to vendor recommendations to achieve comparable count rates to step and shoot acquisition. As an example, 2 minutes/bed translates to a velocity of 0.7 mm/s on a 3-ring PET/CT scanner and 1.1 mm/s on a 4-ring PET/CT scanner. At 5 minutes/bed position, the equivalent bed velocity is 0.3 for a 3-ring PET/CT scanner and 0.4 for a 4-ring PET/CT scanner.

As new technology becomes available, it is important that acquisition parameters are implemented to ensure at least equivalent, if not superior, measurable image quality and output metrics.

Whole body acquisitions can be in either 2- or 3- dimensional mode with attenuation correction, but a consistent method should be chosen for all serial scanning of an individual subject throughout the trial.

A relationship has been described between applied FDG dose, acquisition time per bed position, percentage bed overlap and scanning mode (2D, 3D) in order to harmonize image quality (and avoid bias in quantification) [3, 11]. Using this relationship these parameters are directly linked, e.g. a higher FDG dose can be offset by shorter acquisition times per bed position etc.

Acceptable: All serial scans on any individual subject must be performed on the same previously qualified scanner for each time point if quantitative results are to be used for primary or secondary trial endpoints. If a site has more than one scanner of the same model with the same software version and those scanners have both been previously qualified and both scanners also have been previously demonstrated to be equivalent by periodic quality assurance testing, the serial scans could be performed on any of these equivalent scanners. If a subject has already been injected with the FDG dose and the previously used scanner is not available, a different previously qualified scanner may be used; but this should be noted on the case report form. This may result in restriction of data use to qualitative data only.

If there has been a software version upgrade and pre- and post-upgrade quality assurance testing demonstrates equivalency, this is tantamount to using the same scanner. If there is difference in scanner performance after the software upgrade, this should be noted on the applicable case report forms. This may result in restriction of data use to qualitative data only. All serial scans on the same subject should use identical transmission and emission scanning techniques for all time points.

While there may be variance based on type of scanner, scanning algorithm, model, and software version, the following guidelines are meant to assist each site in achieving the desired data quality as specified in Sections 5.2, 7.1.3, and 12.1.1. Therefore, the determination of the exact scanning acquisition parameters should be guided by the following considerations and activities.

For a dose of 5 MBq/kg or higher (370 MBq or more for a 75 kg patient) the minimal time per bed position using the manufacturers' recommended bed overlap specifications. The time per bed position should be at least 2 mins for 3D systems showing ≥50% bed overlap and at least 3 min for 3D systems showing <50% bed overlap. Time per bed position should be modified inversely proportional to alteration in injected dose per body weight within the limits of the scanner performance as determined by the manufacturer or an appropriately qualified independent standard-setting organization or peer-reviewed publication. There are a wide range of PET systems now available, including time-of-flight, that may change these recommendations. Please see the EANM guidelines for additional recommendations [9].

Commented [MMG2]: This sentence was added after discussion within the writing group. Because of the wide range of PET-CT systems that are now available, and because of the recent introduction of time-of-flight systems, it is impossible to be prescriptive about the optimum time per bed position.

For 2D systems these times per bed should be at least 1.5 times longer for the same injected dose based on body weight. Time per bed position may be modified inversely proportional to alteration in injected dose per body weight within the limits of the scanner performance as determined by the manufacturer or an appropriately qualified independent standard-setting organization or peer-reviewed publication.

In general, increased scan time per bed position will improve the SNR and thus it may be important to increase scan time when quantitative metrics are used towards a primary endpoint.

Whatever scan acquisition parameters are determined on the basis of the recommendations (Acceptable, Target, and Ideal) in this document, efforts should be made to maintain consistency throughout the course of the clinical trial allowing for optional adjustments based on body weight. Specifically, when scan acquisition parameters are determined by quality assessment and control procedures performed for site qualification, those parameters should be implemented for all subjects and all time points, with subject-specific adjustments only as specified and allowed by the imaging protocol embedded within the clinical trial documents. This may require periodic measurement of quality assessment and control parameters and potential subsequent adjustments to scan acquisition parameters after upgrades and major service. All such quality assessment and control procedures should be documented and any resultant adjustments to scan acquisition parameters should also be documented.

Target: Image noise levels are measured using an anthropomorphic phantom (e.g. NEMA, ACR, SNMMI, EANM) with a uniform area to assess image 'noise' by means of the coefficient of variation (COV), which is expressed as a percentage and is defined as COV = (SD / Mean) x 100, for the voxel values within a specified volume of interest (VOI).

The phantom should be filled such that the activity concentration in the uniform area is (approximately 0.1 to 0.2 uC/ml), similar to the expected average normal tissue concentration at the time of imaging in an average weight (70-80 kg) subject in combination with the intended FDG dosage. The phantom should be scanned using the minimal time per bed specified in the trial protocol or using the routinely applied time per bed in the local clinical setting. Moreover, image reconstruction methods and settings should equal those specified in the trial protocol or equal those routinely applied in the local clinical setting.

A volume of interest (VOI) should be positioned entirely within the phantom's uniform area and as much as possible centrally located within the phantom. The VOI should be a cubic or rectangular volume, with the length of each side as close as possible to, but no less than 3 cm. A sphere measuring no less than 3 cm. in diameter may also be used as the VOI on systems that

have the capability to accommodate this strategy. The COV of the voxel values thus determined should be recorded and should also be below 15%. Measurement of noise is encouraged, but is still considered exploratory.

Ideal: Using the methods described immediately above, the phantom should be scanned at the proposed time per bed position and reconstructed using the acceptable reconstruction methods and settings (e.g. minimal and/or harmonized resolution criteria). The COV within the VOI should be calculated and should yield a COV of 10% or better. If the ideal COV is not achieved, the time per bed position could be increased so as to achieve the desired COV.

7.2.1. Subject Positioning

During PET-CT, subjects should be positioned in the center of the field of view (FOV), preferably with the subjects' arms to be positioned overhead (to minimize beam hardening and FOV truncation artifacts). Alternatively, the arms can be positioned along the side for head and neck imaging (for two-step procedure – see section 7.1.1). Subjects may be unable to maintain arms above head for the examination, in which case protocol specific handling needs to be defined. Arm positioning in a particular subject should be consistent as possible across all time points.

If PET-CT data are used for radiation planning, the examination should be carried out in the radiation position using the same dedicated radio-opaque positioning devices as used in the radiotherapy department. Support devices, under the back and/or the legs, may be used to enable the subject to comfortably maintain his/her position throughout the exam [28].

7.2.2. Instructions to Subject during Acquisition

The diagnostic CT is usually performed in maximal inspiration breath-hold which could result in image artifacts due to mis-registration of the lung-liver interface between emission and CT images if the diagnostic CT is being used for attenuation correction (i.e., there is only one CT scan performed for both diagnosis and attenuation correction which is not the UPICT recommended method per section 7.1.1). Therefore, the CT acquisition for attenuation correction should be done with shallow breathing without regard to the CT technology used (acceptable / target / ideal).

7.2.3. Timing / Triggers {*This section intentionally omitted*.}

7.2.4. Model-Specific Parameters

The vendor model-specific and software version-specific parameters that would reproducibly produce image data meeting the requirements as stated in Section 7.1., while also complying with the radiation dosimetry as specified in Section 12 and 13 is not known at this time. Optimally, the vendors will, over time, produce such operating instructions for some if not all of their platforms. For the present, this document specifies certain performance criteria and image quality specifications that must be met as described elsewhere in this section.

7.3. Imaging Data Reconstruction

- PET emission data must be corrected for geometrical response and detector efficiency (normalization), system dead time, random coincidences, scatter and attenuation [3, 11, 28].
- Data acquired in the 3D mode can be reconstructed directly using a 3D reconstruction algorithm or re-binned into 2D data and subsequently be reconstructed with a 2D reconstruction algorithm.
- Iterative reconstruction algorithms are current standard for PET (rather than filtered back projection), and should be used to reconstruct all PET images.
- Reconstructions should be performed with and without attenuation correction.
- Scanners must be properly normalized and calibrated to ensure uniformity and accuracy of SUV measurements within the limits of the spatial resolution
- Standardization of reconstruction performance is necessary to obtain comparable resolution and SUV recoveries across the same subject and inter-subject across sites. This has not yet been achieved, but is actively being addressed by the major PET manufacturers.

7.3.1. Model-Specific Parameters

Acceptable: The current acceptable practice is to provide general reconstruction guidelines and allow individual sites to choose the specific parameters used for their particular scanner model/version, based in part on current clinical practice. If this approach is used, the parameters should be reviewed for appropriateness and consistency and the resulting image quality should be assessed with phantom imaging performed as part of the PET/CT scanner qualification.

Target/Ideal: If warranted by the particular trial endpoints (and specifically if an endpoint is based on absolute quantitative PET measures), acquisition and reconstruction parameters for each specific scanner model/version should be tailored to achieve comparable performance (i.e., harmonization across platforms and sites) in terms of spatial resolution or SUV contrast recovery and noise.

7.3.2. Archival Requirements for Reconstructed Imaging Data -- See 11.4.

7.3.3. Quality Control -- See 12.4.

8. Image Post-processing

8.1. Input Data to Be Used

Input data can be either Reconstructed Data, or Post-Processed Image Data as defined below.

8.1.1. Definitions

Raw Data: This is an ambiguous term as it can refer to scanner raw data (i.e., sinograms or list-mode) or image raw data. This term should not be used.

Raw Projection Data: This term refers to the data as acquired by the scanner before reconstruction (i.e., sinograms or list-mode). When this term is used, the user should specify the exact type of Raw Projection Data.

Reconstructed Image Data: This is the image data exactly as produced by the reconstruction process on the PET or PET/CT scanner, i.e., a stack of DICOM slices/files constituting a PET image volume with no processing other than that occurring during image reconstruction. This is always a stack of DICOM slices/files constituting a PET image volume that can be analyzed on one or more of the following: PET scanner console, PET image display workstation, PACS system, etc.

Post-Processed Image Data: An image that has been transformed after reconstruction in some manner, including but not limited to: smoothing, sharpening, image zoom, rotation/translation, resampling, interpolation, slice averaging, MIP, etc. This is typically a stack of DICOM slices/files constituting a PET image volume that can still be analyzed on one or more of the following: PET scanner console, PET image display workstation, PACS system, etc.

Secondary Image: This is an ambiguous term as it can refer to either Post-Processed Image Data or a DICOM secondary capture image (akin to a photograph). This term should not be used. Instead please see Post-Processed Image Data above.

8.2. Methods to Be Used

After data collection and image reconstruction as detailed in Section 7, Reconstructed Image Data (PET images) are generated that meet the image characteristics defined the by the trial.

For both visualization/interpretation and quantification, no unintended additional image processing (interpolation, re-binning, reorientation, zooming etc.) should be applied to the originally reconstructed PET data.

8.2.1. Definitions

Image Processing: Transformations applied to an entire image or a region of an image. These transformations include, but are not limited to: smoothing, resolution recovery, image zoom, rotation/translation, re-sampling, interpolation, slice averaging, de-identification, etc. The output of this process is itself an image, often intended for visual or quantitative analysis.

8.2.2. Processing affecting quantification

Acceptable: Image Post-Processing methods and parameters that are used should be recorded and applied to all images in a consistent manner following methods specified in the clinical trial. For example all images might be smoothed to the same overall resolution and/or reconstructed with the same voxel size (or in a defined range of voxel sizes). Quantitation should be applied consistently across all time points and all subjects within a given site.

The originally reconstructed PET data set should always be preserved. In case processed PET datasets are required, they should be saved as separate secondary datasets.

Target: No Image Post-Processing is used for quantitation and all analyses are applied to the Reconstructed Image Data. Post-Processed Data may be used for visualization and to facilitate identifying the ROI / VOI. However, the underlying Reconstructed Image Data should be used for all quantitative purposes. The ROI / VOI derived from the Post-Processing should be transferred to the Reconstructed Image Data for quantitation. Quantitation should be applied consistently across all time points and all subjects within a given site.

Ideal: No Image Processing is used for quantitation. Instead the analysis software for ROIs and VOIs always applies the analysis to the Reconstructed Image Data, regardless of the appearance of the image on the display station (which may be Post-processed). This is also a component of the QIBA FDG-PET Profile [10]. The Ideal level of performance is equivalent to the Target level of performance, but in addition to being applied consistently across all time points and all subjects within a given site the consistency is also across all subjects, all time points, and all sites within a given trial.

8.2.3. Processing affecting visualization

Additional image processing may be performed for specific applications or use cases. For visualization most of the image viewing software or platforms will 'automatically' apply some kind of image interpolation (on screen) and image zoom to enhance visual image quality, i.e., almost all viewing and data analysis SW application will perform online image interpolation while displaying PET images on screen. Additional image processing may be applied upon user input, such as zooming, re-binning, reorientation, adjustment of slice thickness or summing of slices and image filtering. When automatic interpolation is applied, it would be desirable that the user has accessibility to replicated zoomed image data at its original matrix size.

Acceptable/Target/Ideal: For visual inspection/interpretation of PET/CT data the by the viewing software or platform default online interpolation and zooming may be used. In addition, so-called maximum intensity projections (MIP) may be generated as they may facilitate localization and detection of lesions. Additional processing, such as zooming, re-binning, reorientation and filtering may be applied upon user request only. User should be able to manipulate color scale settings (window/level and color table). It should always be possible to revert to the default orientation, zoom and bin-size (preferably a 'revert to default' button is available).

8.2.4. Image de-identification (See also Section 11.2)

Acceptable: If images are de-identified to remove PHI, no information that affects quantitation should be removed.

Target/Ideal: Only the minimal required PHI should be removed; i.e., all information that is not required to be removed should be retained.

8.3. Required Characteristics of Resulting Data

Acceptable: After visual post-processing is completed, the original data subjected to the post-processing must be retained in its original state. The transformation between the post-processed and original data must be described so as to allow subsequent reproduction by a third party. Any annotations and/or mark-ups performed on the post-processed dataset must be transformed to a copy of the original dataset (but still leaving one copy of the original dataset without alteration).

After PHI is removed, all information that affects quantitation should remain intact and unchanged.

8.4. Platform-specific Instructions

Currently there are no specific instructions that have been compiled for various platforms. Post-processing should be performed in accordance with vendor recommendations for the given model and/or specific user manuals.

8.5. Archival Requirements -- See 11.5.

8.6. Quality Control -- See 12.5.

9. Image Analysis

For quantitation to be most robustly applied, images must meet the image acquisition guidelines as stated within the UPICT Protocol, including, but not limited to, similar tracer uptake times (see Section 5.3), same scanner and reconstruction algorithm (see Section 7.3) and similar injected dose (see Section 5.2). Additionally, the same software and workstation model and version should be used for a given subject across all time points (and for central analysis for all sites and all subjects and all time points) for the analyses described in this section. Stability and acceptability guidelines have been articulated in the PERCIST 1.0 guidelines [31].

Image analysis and interpretation also presumes that the image datasets to be used are reconstructed and attenuation corrected as per 7.3 of this UPICT Protocol.

9.1. Input Data to Be Used and Covariates Necessary for Analysis

Image quantitation is typically performed by determining a Standardized Uptake Value (SUV) in tumor and, ideally, in a reference normal organ. The SUV measure to be utilized needs to be specified for each protocol and needs to be used consistently at all sites and across all subjects and all time points for all lesion measurements. The accuracy of the SUV obtained from each workstation should be verified as specified in section 9.1.3.

9.1.1. The SUV Statistic

Nomenclature relevant to the SUV statistic shall be defined to address the (1) subject relevant versus (2) statistical sampling relevant issues. Regardless the SUV statistic(s) used, it is recommended that the SUV value is recorded at least to the tenths place (e.g. 4.7) whether used as an absolute value or as a change metric. As an exploratory metric, it is

suggested that some measure (e.g., SD) of heterogeneity within measured multi-voxel VOIs be expressed along with the SUV metric (e.g., 4.7 ± 0.2). However, it should be recognized that the utility of reporting this variance in is unknown at this time and is likely highly dependent on the standardization of the imaging and reconstruction processes.

9.1.1.1. Subject indices (bw, lbm, bsa, other)

Target/acceptable is SUV reporting with inclusion of measurement and reporting of subject height and weight (see separate section 4.2.2) and reporting to allow for other normalizations.

The subject relevant issue is whether to use body weight (bw), lean body mass (lbm) or body surface area (bsa).

- SUL = SUVlbm = reference to lean body mass
- SUV = SUVbw = reference to body weight
- SUVbsa = reference to body surface area (rarely used)

From the SNM GHS [14], there was consensus that SUV normalized to lean body mass (SUL) is an appealing concept for correcting the radiotracer distribution based on differences in body habitus in order to obtain absolute values and changes. It was acknowledged that the requirement of SUL may be limiting at this time due to either vendor platform software limitations, or limitations in the formula for characterizing the obese patient population.

Lean-body-mass normalization has most commonly been done using the James formulas [32] (weight as Kg, height as cm):

- LBM(male) = 1.1 x Weight - 128 x $Weight^2/Height^2$
- LBM(female) = 1.07 x Weight – 148 x $Weight^2/Height^2$

While the James formulas perform well at body weights below approximately 100 Kg for females and 140 Kg for males, the formulas are increasingly inaccurate at higher weights [33]. These formulas have been incorporated into several software systems for viewing and analyzing PET images. The results are acceptable as long as the weight limitations are respected.

However in the future, when lean-body-mass normalization is used for SUV calculation, the target recommendation is to use the formulas developed by Janmahasatian [34] which are much more plausible at high patient weights (weight as Kg, body mass index (BMI) as Kg/M^2):

- LBM(male) = 9270 x Weight / (6680 + 216 x BMI)
- LBM(female) = 9270 x Weight / (8780 + 244 x BMI)

An incorrect, alternative form for males has sometimes been used, which can be traced back to an article by Morgan and Bray [35] in which the formula presented by James is likely misquoted, using 120 instead of 128 as a coefficient. This form was mentioned, but not used, in an article by Sugawara et al. [36], as a method for

Commented [MMG3]: The discussion regarding the optimum formula for lean body mass was expanded to include the concept of using the Janmahasatian formulas in the future, both because the results are far more plausible and to be in harmony with the EANM guidelines.

LBM normalization of SUV calculations, with subsequent adoption by some practitioners. However the pharmacology community does not use the alternative version [37].

There are also continuing efforts to come up with improved methods for estimating LBM, including direct measurement on a per-patient basis using CT. [38] However, as noted in Appendix H of the QIBA FDG-PET/CT Profile [10], the different methods provide estimates of LBM typically have unknown levels of bias and variance.

9.1.2. Statistical sampling – including report-out values: ROIs and VOIs

Each of the SUV statistics defined above may be measured by one of three statistical sampling methods. That is the SUL, SUV, and SUVbsa may each be measured using a single voxel measure (max) or multi-voxel measures (mean or peak). There are known issues with the use of the SUVmax in the presence of low counts, which result in positive bias [39], specifically there is an upward bias of the single voxel SUV max at low count rates. In addition, multiple voxel methods have shown improved repeatability [15, 39, 40]. Despite these issues, the SUVmax has demonstrated utility as a prognostic and predictive indicator in both clinical use and research studies, even though it may not be as reproducible from study to study as the SUV of larger regions. The following discussion (and the remainder of Sections 9 and 10) will use SUV as the generic example. However, the discussions are generally applicable to SUL and SUVbsa (when appropriate and necessary discussion differentiating among these statistics will be included in various sections of this document).

- SUVmax = single voxel (most FDG-avid voxel in tumor ROI)
- SUVmean = mean SUV value for ROI with more than one voxel
- SUVpeak = subcategory of SUVmean where volume (SUVpeak-3D) or area (SUVpeak-2D) is defined specifically.

In PERCIST, the SULpeak is a 3D ROI obtained from a 1 cc volume sphere (measuring approximately 1.2 cm in diameter) and defines the most metabolically active 1 cc volume in a tumor. An approximation of the SUL peak can be the value obtained by measuring the SUVpeak of an area which is 1.2 cm in diameter and which usually subtends only a single slice, but which might also be defined on multiple (most usually three) slices (for further discussion on the methods to be used for defining the 3D volume and the 2D area, please see Section 9.2) ACRIN defines the 2D SUVpeak as a circular ROI centered on the SUVmax with a 0.75-1.75 cm diameter (1.0 cm is preferred) [31]. Some PET workstations do not have automated methods to define the SUV peak. There are alternate approaches for determining the region to be used for the SUVpeak metric. One involves moving the VOI/ROI throughout the tumor and measuring multiple SUVpeaks (one for each VOI/ROI) until the highest intratumoral SUVpeak measurement is located. Another involves locating the SUVmax and then centering the SUVpeak VOI/ROI on the SUVmax pixel. However, this method may not result in measuring the most FDG-avid portion of the tumor. An automated search mechanism to find the most FDG-avid SUVpeak has been developed as a computer code in some systems. It is often, though not always, the case that SUVpeak is centered on the SUVmax pixel in a tumor. It would be ideal to achieve consistency in the peak method that is used. However, it is unclear at this time which method is optimal.

All references indicate that SUVmax (maximum voxel value or most FDG-avid voxel) is required for each lesion that is reported as specified in the study protocol and/or considered clinically relevant.

Multiple references also indicate that SUVmean of the VOI/ROI obtained be reported [2, 3, 41]. The SUVpeak equals the SUVmean only when the VOI is a sphere with a specified diameter, which is also indicated as a reportable statistic and the SUVpeak is the most intense region of the tumor. PERCIST requires the use of SULpeak [31]. The SUV mean may be operator and ROI placement dependent if defined manually. While it has been used in many studies, it is not required by PERCIST as is SUV max. More objective methods are preferred for segmenting the tumor to define SUV mean (see sec 9.2).

Nearly all PET systems will allow determination and reporting of a single voxel SUVmax. However, several reproducibility studies have shown somewhat greater variance for single voxel measurements (SUVmax) on test/re-test than for somewhat larger regions of interest (SUVmean) [15]. Newer PET scanners offer PET reconstructions including matrix sizes of 256 x 256 and larger and slice thicknesses in the 1-2 mm range. These single voxels are much smaller than the single voxels used in earlier determinations of PET precision and are more subject to noise related variance. At low count levels these single voxel measurements are subject to systematic errors including possible overestimation of SUVmax as compared with truth. In addition, point spread function/resolution recovery methods have been implemented which may variably drive single voxel quantification.

While these methods have been used to improve lesion detection, there are changes in quantitative values that may impact response assessment. At this time, It is preferred that studies with quantitative response assessment not use resolution recovery methods due to the unknown impact and lack of standardization. For this reason, while single voxel values can be reported and are typically highly correlated (though higher) with an SUVmean from larger VOI (such as the 1.2 cm diameter volume recommended in PERCIST, SUVpeak), caution must be given to modest changes in values in single voxel SUVmax from test to test, especially in newer PET scanners with short acquisitions, large matrix sizes, low injected tracer doses and thin slice thicknesses (resulting in small voxels). Most contemporary PET workstations allow for determination of a VOI of a fixed volume larger than a single voxel. At present, variance of the SUV in a larger VOI is not reported, but it may be explored.

The optimal method of assessing a biologically relevant tumor response may vary depending on the tumor type, therapy, and timing of scans vs. the therapy, and is not yet fully resolved. Furthermore, the underlying tasks of choosing and prioritizing the optimal statistical metric to use and the optimal methodology to define lesion VOI/ROI (section 9.2) is challenging given the lack of rigorous comparative studies to date on which to rely. It is clear that the differing metrics are strongly correlated with one another. Methods with a single voxel are statistically more variable than those with slightly larger numbers of voxels included; meaning that changes in single voxel SUV measure (i.e., SUL, SUV, SUVbsa) between studies may have to be larger to be statistically different. Intuitively, the most accurate representation of a lesions cellular tumor burden should include a combination of tumor burden volume and the metabolic activity of that burden as proposed with the Total Lesion Glycolysis (TLG) [42]. For very small tumors, the SUVpeak values may include some

tissue that is non-tumor, lowering apparent tumor activity. It is also possible tumor volume from PET may be informative.

Note that by combining strategies of body habitus normalization and ROI peak averaging using the PERCIST example of SULpeak, this is an SUV measurement using lbm as patient size normalization and mean value of specific size (1.2cm diameter sphere) VOI/ROI as statistical sampling method. Furthermore, SUVpeak can be provided which uses bw as subject distribution "unit" and mean value of specific size VOI/ROI as statistical sampling method.

Acceptable: SUVmax (normalized by body weight or lean body mass) - single voxel (must specify and should be the same across all subjects and time points); x, y, and z dimensions of a single voxel should be known and recorded (e.g. within the DICOM header). Input parameters for calculating SUV should be recorded (section 9.1.2).

Target: SUVpeak in addition to SUVmax (must specify and should be the same across all subjects and time points). For discussion of how partial or fractional pixel / voxel data could and should be managed, see Section 9.2.2.

Ideal: In addition to recording the Target metrics, additional metrics for body habitus correction and/or voxel averaging should be included such as the SULpeak (SULpeak-3D more desirable than SULpeak-2D) and SULmax- both in the most FDG-avid region of each particular target tumor should be captured - size of single pixel should be known

Exploratory: it is recommended but not required to supplement Ideal, Target, and Acceptable performance with an exploratory measures of Total Lesion Glycolytic (TLG) activity [39] and Metabolic Tumor Volume (MTV)

9.1.3. Covariate inputs (e.g. glucose uptake time, height, weight, FDG-dose)
Please see Section 4.2.2 on obtaining and recording covariate inputs and Section 10.2.1.5.1 regarding glucose correction.

9.2. Methods to Be Used

9.2.1. Methodology for defining ROI/VOI

ROI (or VOI) tool to be utilized to define either fixed symmetrical size object or lesion constraint condition and strategy to define edge detection needs to be prescribed. Note that the methods for extracting metrics from ROI/VOIs are described above in section 9.1. To follow is a catalogue of potential strategies, but the UPICT Protocol does not stipulate any one as preferred. However, the trial design should stipulate which of the strategies is to be used uniformly across all subjects and time points during the course of the trial. These strategies can be summarized as below:

Manual: Requires the intervention of an expert reader to define anatomic and/or metabolic ROI/VOIs. While this method does not represent ground truth it may be used as a standard for the apparent tumor boundaries, it is observer dependent and may have substantial inter- and intra-reader variability. 3D manual approaches require defining ROIs

on multiple planes to generate VOIs. Likewise, a 3D measurement such as SUVmax requires evaluating multiple 2D ROIs to identify the plane containing the maximum SUV within the tumor volume. Shapes can either be irregular polygons or fixed geometric shapes such as circles, rectangles, etc.

Semi-automated: Requires some user intervention such as defining target lesions or masking neighboring healthy structures with physiologic FDG-uptake and uses computer algorithms to define tumor boundaries. A common approach is to use either a pre-defined or user-defined relative threshold based on the maximum value (e.g. 70% of SUVmax). Another approach is to use an absolute threshold (e.g. SUV liver mean + 2SD). More sophisticated approaches have also been implemented such as using gradient-based segmentation.

Automated: Requires no user intervention and is fully automated. However, algorithms must be validated against ROI/VOIs defined by expert readers.

By way of an example, the threshold for definition of an evaluable lesion for tumor volume articulated by PERCIST is mean liver SUL in a 3 cm. diameter sphere in the right lobe of the liver + 2 SD of liver noise. This threshold is defined at baseline so that lesions can be "hot enough" to have a measurable decline in F18 activity on subsequent studies with therapy. For relative threshold as the constraint definition, SNM GHS notes that tumor ROI's reflecting the metabolic volume of the tumors are desirable. For simplicity, volumes based on a 70% threshold of the peak tumor SUV should be produced. These approaches are viewed as exploratory, but recognize that tumor volume may provide useful information beyond that of the peak or max SUV in a tumor.

9.2.2. Geometric issues (e.g. handling partial pixel/voxel)

The SNM GHS suggested that appropriate use of partial pixel values to secure a 1.2cm diameter (≈1 cc volume) ROI was appropriate and desirable, since standard pixel sizes would not allow selection of a 1 cc volume precisely in most cases [14].

Acceptable: Any regular 2D area for peak activity measurement (e.g. SUVpeak-2D) ROI would be defined as a circular ROI on a single axial slice with a diameter of 1.2 cm within the limits of the voxel size (with a minimum diameter of 3 voxels without using partial voxels). It is also acceptable to use a 1.2 cm circular ROI with interpolated voxel values.

Target: Any regular 3D volume for peak activity measurement (e.g., SUVpeak-3D) VOI would be defined as an isotropic spherical VOI with a diameter of 1.2 cm within the limits of the voxel size (with a minimum diameter of 3 voxels without using partial voxels).

Ideal: Any regular 3D volume for peak activity measurement (e.g. SUVpeak-3D) VOI would be defined as an isotropic spherical VOI with a diameter of 1.2 cm (achieved using interpolated voxel values).

Exploratory: For irregular VOI (TLG, MTV) no single method is specified as Ideal or Target. However, Acceptable performance of this exploratory metric is defined as specifying which

method is used and using the same method consistently across all time points for all subjects and sites, and providing the data as stated in Section 9.1.

9.3. Required Characteristics of Resulting Data

9.3.1. Tumor assessment – See Sections 9.1, 9.2, and 10.

9.3.2. Internal normalization / Comparator tissue(s)

The stability of normal tissue SUV (e.g. liver, blood pool) in tests performed at differing times in the same patient is considered to be a reasonable and practical indicator of the use of similar techniques of performance of PET (see 12.3) when quantitative FDG-PET/CT is used as a primary or secondary endpoint [43]. Such stability can suggest it appropriate to use the tumor SUV data for response assessment. Measurement of the normal liver mean was suggested using a 3 cm diameter spherical VOI that should be reported at each time point. An alternate method is use of blood pool activity (especially if the liver is adversely affected by metastatic disease) (as described separately -reference section 10.2.1.1.1).

It is possible that a subject's liver SUV may change during the course of the trial (perhaps as a consequence of disease progression or the therapeutic intervention). The study protocol should specify how quantitative measurements in subjects with "out of range" liver (blood pool) SUL measurements will be managed. One potential mechanism would be to analyze the data both including and excluding subjects with "out of range" liver (blood pool) SUL measurements.

Acceptable: SUV of the liver and/or blood pool should be reported for all subjects and all time points. Large deviations in SUVs between the baseline and follow-up time points should be investigated for technical errors (e.g. incorrect dose or calibration issues).

Target: If the SUV of the liver and/or blood pool are not within 30% of the comparator (either baseline or immediate previous as dictated by the study protocol) study then the data receive additional level of review and scrutiny to determine if it should be included in the study. PERCIST proposed the following: Normal liver SUL must be within 20% (and 0.3 SUL mean units) for baseline and follow-up study to be assessable. If liver is abnormal, blood pool SUL must be within 20% (and 0.3 SUL mean units) for baseline and follow-up study to be assessable.

Ideal: Unknown
Exploratory: The ratio of tumor SULpeak to liver (blood pool) SULmean could be reported as an exploratory metric to correct for global variations.

Liver (or blood pool) SULmean and SD are important to report, but not a full substitute for quality control (see Section 9.5.1.2). Liver (or blood pool if liver is replaced with disease) ROI/VOIs are considered a reasonable method to assess noise, although acceptable noise level in PET has not yet been determined.

Acceptable: Qualitative visual assessment should be performed to confirm the overall image quality and noise are acceptable.

Target: SD of liver or blood pool recorded at baseline and all subsequent time points.

Ideal: Normal tissue SD such as liver or blood pool would ideally be used to assess image noise and define quality control procedures.

9.4. Platform-specific instructions {*This section intentionally omitted*.}

9.5. Archival Requirements

Any annotations and/or mark-ups performed during post-processing and/or analysis must be transformed to a copy of the original dataset (but still leaving one copy of the original dataset without alteration); also please see 11.6.

9.6. Quality Control – (Also see 12.6.)

9.6.1. Statistical Quality of measurement(s) (e.g. noise)

Quality control of the required inputs (imaging data acquisition and reconstruction and covariates) has been described elsewhere in this document and must be satisfied prior to analysis and interpretation. Additional QC metrics should include:

9.6.1.1. Subjective assessment of image quality.
For example, movement or mis-registration can lead to invalid AC, poor quality / unreliable quantitative data. Some images may be too poor in quality (e.g. inadequate counts per field) to quantify. All necessary data available to determine if quality is acceptable or not (e.g. both AC and non-AC images) should be generated routinely and must be available. Specific sources of degradation in quality that should be assessed include, but are not limited to:

- Artifacts secondary to implants in area of concern
- Patient motion
- Extraneous activity (e.g., IV tubing or urine) in field.
- Extravasation of FDG

The output of this subjective QC assessment must include the judgments to whether the study, despite artifacts, still has utility in analysis (e.g. quantitative, semi-quantitative, and/or qualitative).

9.6.1.2. Objective Assessment

Ideal: Use of a digital reference object is necessary to assess the performance characteristics (e.g. accuracy, precision, etc.) of the software tool, the user interface, and the "user" during the SUV determination workflow including, but not limited to, the determination of the most FDG-avid pixel / voxel and the creation of the standardized ROI / VOI.

Acceptable / Target: Document the workstation and software models and versions used and ensure that for each subject the same workstation and

software model and version is used across all time points; should hardware and/or software upgrades occur during the course of the trial, testing should verify the comparability of quantitative metrics used in the trial (with comparability defined by the specifications in the clinical trial documentation) also see Section 12.1.1.

Document that the selected parameters used for analysis were achieved in actual practice. All workstations and software tools should have gone through validation by the manufacturer with approval by the appropriate regulatory body(ies) or the validation should be publically and transparently available. The trial should include specific QC tasks to ensure QC of the users with documentation at the time of site qualification and periodically during the trial.

10. Image Interpretation

10.1. Input Data to Be Used {*This section intentionally omitted.*}

10.2. Methods to Be Used

The points listed serve to take the input data and then:

a) *discriminate* - qualify as either target or non-target lesion
b) *compare* - to baseline
c) *derive*- use combination of target / non-target / presence/absence of new disease to *describe, stratify, and potentially classify or categorize into discrete classifications* –

into response assessment category (responder, stable, progressive disease) to obtain Output data (which could also include SUL data of each lesion) from which an Interpretation (Section 10.3-Required Characteristics of Resulting Data) can be rendered (with incorporation of QC check). There are overlap issues (to Baseline and On-study time points), but there are also time-point specific issues which discriminate Baseline from On-study.

10.2.1. Baseline Time Point Evaluation

10.2.1.1. Qualification of Target Lesions

Important definitions
PD: Persistent Disease
PR: Partial Response
PMD: Persistent Disease
PMR: Partial Response
SMD: Stable Metabolic Disease
CMR: Complete Metabolic Response

While target lesions require the most FDG-avidity, If the lesion cannot be reliably be measured on PET due to, for example, artifacts from nearby intense F18 containing structures (like the bladder), then an alternative the next most FDG-avid measurable lesion can be quantified. Similarly, if the most FDG-avid lesion is in a region where the quality of quantitation is suspect perhaps due to motion or

attenuation artifacts (e.g. at the diaphragm/liver interface, or in the neck under the circumstance that the head has moved) then (an) alternative lesion(s) can be chosen, ideally nearly as intense in activity. The less easily measurable lesion would be a non-target lesion and would still be assessed for disappearance in the case of possible PR or clear increase in activity in the case of PD. While PERCIST does not require a lesion to be measurable by CT or anatomic measures when choosing (a) target lesion(s), if two lesions are of similar FDG avidity (i.e., within 10-15% of one another), then the lesion which is more easily measurable anatomically might be preferable for analysis. Details are enumerated below.

Commented [MMG4]: We need to define PR, PD, PMD, PMR, CMR, SMD see above. Should they also go on page 67?

10.2.1.1.1. Minimum metabolic threshold

If using a single lesion paradigm for change assessment, the most FDG-avid lesion should be selected. However, if this lesion cannot be reliably measured on PET due to, for example, artifacts from nearby intense F18 containing structures (like the bladder), then the next most FDG-avid lesion should be measured. Similarly, if the candidate target lesion is in a region where the quality of quantitation is suspect, perhaps due to motion or attenuation artifacts (e.g. at the diaphragm/liver interface, or in the neck under the circumstance that the head has moved), then (an) alternative lesion(s) can be chosen, ideally nearly as intense in activity.

If a multiple target lesion paradigm for change assessment is used, then the aforementioned considerations for target lesion selection should also be applied. In either case (single or multiple target lesion selection), the less easily measurable lesion(s) would be non-target lesion(s) and would still be assessed for disappearance in the case of possible PR or clear increase in activity in the case of PD. While PERCIST does not require a lesion to be measurable by CT or anatomic measures when choosing (a) target lesion(s), if two lesions are of similar FDG avidity (i.e., within 10-15% of one another), then the lesion which is more easily measurable anatomically might be preferable for analysis. PERCIST proposes 1.5 x liver mean SUL (3 cm diameter spherical ROI in the right lobe of normal liver) + 2 X SD of liver noise as the minimum target lesion threshold at baseline. If the liver is not in the field of view or is abnormal to a degree that normal liver cannot be assessed, then the alternate comparator is to use a minimum threshold level of 2 times SUL mean of blood pool in a 3D object defined as a 1-cm diameter ROI in descending thoracic aorta extended over 2-cms tracking the long axis of the aorta; or by making this measurement in multiple 2D 1-cm diameter ROIs extending sequentially over 2-cm of the descending aorta. If the descending aorta is not evaluable a VOI of the same volume should be measured from elsewhere in the thoracic aorta.

Given the absence of knowledge the general guidance is suggested below:

Acceptable: A minimum FDG-avidity is required and should be specified in the clinical trial protocol. This can be determined by either a subject-specific threshold as proposed with PERCIST [31] or as a general cutoff. For a general cutoff, an SUVmax of 4 is suggested for all target lesions, although in some settings a lower minimum SUVmax may be acceptable, such as in the lung or breast.

Target/Ideal: The ideal minimum threshold above background is not known. Components of the ideal threshold could include both the mean and standard deviation of the SUV of a normal reference tissue.

10.2.1.1.2. Influence of anatomic measurability of lesion size; including reportability of lesion anatomic size

In PERCIST 1.0, lesions selected as target lesions on the basis of meeting minimum metabolic activity thresholds as defined above (Section 10.2.1.1.1) need not meet minimum size requirements; although if multiple lesions with similar FDG activity are present, the most FDG-avid anatomically measurable lesion(s) are preferable to FDG-avid lesion(s) that are not anatomically measurable. This may be more valid for lesions that are markedly FDG-avid than for lesions that show relatively low-level FDG activity. Therefore by extension for lesions that have less FDG avidity, it may be reasonable to include a minimum lesion size threshold (or guideline) in addition to other minimum criteria for target lesion qualification.

This is especially important for small lesions in anatomic areas subject to artifact from motion (e.g., lung base or hepatic dome) or for lesions difficult to separate from contiguous normal tissues showing metabolic activity (e.g. urinary bladder). The SNM GHS* suggests that tumors should typically be over 2 cm in diameter for target lesion inclusion at baseline, although a lesion meeting the appropriate FDG activity metrics need not meet this anatomic measurement threshold as a mandatory minimum.

Practically, evaluation of lesion size (e.g., longest diameter) may be difficult, especially if no dedicated CT was performed either in conjunction with or within an allowable temporal association with the FDG-PET scan. This may be due to intrinsic lesion characteristics (e.g., infiltrative or CT lesion isodensity to surrounding tissue) or due to the anatomic location of tumor (e.g., bone marrow site). For lesions subject to partial volume effect of

SUV measurement, notably due to anatomic location (e.g. peri-diaghragmatic lesions at either lung base or hepatic dome), a minimum size requirement may also be reasonable.

If multiple candidate target lesions of similar FDG intensity are present, then the chosen target (or targets depending upon response assessment paradigm being used) should be the larger of the lesion(s) also taking into account the reproducibility of lesion measurement based on subjective factors described below (Section 10.2.1.3).

These issues should be addressed prospectively in the clinical trial protocol and protocol-specific guidelines should document whether or not minimum size criteria for target lesion qualifications are used and if so how such size criteria will be used.

Subjective assessment on reproducibility of measurement (e.g., contiguous structures, conglomerate lesions, hypometabolic lesions, fluid collections, etc.)

Given multiple lesions that qualify on the basis of threshold activity and minimum size, priority should be given to those lesions that are measurable in an accurate and reproducible way. Therefore, lesions with a problematic anatomic location or configuration might not be chosen for measurement if there are other lesions that may be measured with more accuracy and reproducibility. If a lesion is not chosen at baseline secondary to difficulty in accurate measurement, but on subsequent scans the lesion is assessed as dominant or progressive then hindsight review may be appropriate. The analysis and interpretation should explain the interscan discrepancy (see section 10.3) and such a lesion may have to be assessed as a "non-target" lesion.

10.2.1.2. Use of Non-target lesions

Non-target lesions can be considered as disease that is quantifiable or disease that is assessable qualitatively but does not meet requirements for target disease. The presence of non-target lesions should be noted; this can be done either by noting the presence/absence of non-target disease or by identifying sites of non-target disease by organ or anatomic location (e.g., liver or abdominal nodes). Non-target disease should be qualitatively evaluated at each time point. Furthermore, changes in the status of the non-target lesions may be noted if only in a qualitative manner (see section 10.2.1.3). However, if a non-target lesion becomes a target lesion on a later scan, hindsight quantitative review may be appropriate. The analysis and interpretation should explain the interscan discrepancy (see section 10.3). Note that in PERCIST, non-target lesion(s) can become target if the lesion increases in intensity beyond the original target lesion, such that the previously defined non-target lesion is the most FDG-avid

lesion on the subsequent scan performed on-study. This would typically be considered disease progression if PERCIST criteria are met.

10.2.1.3. Use of Qualitative lesion assessment
Incorporation of a visual assessment in the analysis and interpretation with documentation in the CRF may have utility especially in certain oncologic conditions (e.g. Cheson criteria in lymphoma).

10.2.1.4. Other Observations and reporting methods

The assessment should include commentary related to false positive and false negative (e.g. disease mimics/variants/QC) activity, as not all foci that meet the preceding criteria may be indicative of disease (e.g. infection, inflammation, fracture, post-radiation changes). Similarly, there may be artifacts that mimic or obscure reportable disease (e.g. metallic orthopedic and/or dental implants). The trial case report forms should include a mechanism for ensuring the capture of these data.

10.2.1.5. Covariate & Normalization Strategies

10.2.1.5.1. What to use and what not to use (e.g. glucose correction)

Glucose normalization (both for SUV and SUL):
Acceptable – Collect glucose data on everyone shortly before radiotracer is injected;
Target – Use a properly calibrated glucometer and collect glucose data;
Ideal – It is not clear yet if corrections for glucose levels enhance the ability of PET to predict treatment response. It is suggested this can be explored prospectively to help determine if the actual corrections of SUV are appropriate / necessary / possible. It is possible the "corrections" may add additional errors to assessments so it is not viewed as appropriate to routinely apply "corrections" in this setting [44].

Correction for the timing of image acquisition relative to the time of FDG injection outside the prescribed window has been suggested by some references. However, this is not universally accepted and is considered exploratory at this time.

10.2.2. On-study Evaluation

10.2.2.1. Strategy dependent upon the analysis and interpretation paradigm

The workflow for the analysis and interpretation of the non-baseline imaging examinations (i.e. “on-study” evaluations) is based on the response assessment paradigm that has been chosen for the specific clinical trial; and therefore the baseline requirements.

A reviewer's approach to performing target lesion inter-time point FDG-PET assessment depends primarily upon the interpretation strategy, distinguished by two considerations:

- By using either one target lesion or up to five target lesions and
- By using the most FDG-avid lesion(s) for each time point versus comparing the same lesion(s) across time points

The imaging review charter should define the approach prospectively. Currently, the literature is not conclusive on which approach best correlates with clinical outcomes. In order to obtain data consistently across multiple studies that can eventually undergo meta-analysis, it is recommended to perform quantitative analysis on up to five of the most metabolically active lesions, to include the most metabolically active lesion at each time point. The details of how to perform this analysis are included in the target lesion section below. The case report form (and subsequent data capture) should be structured in a manner to allow both cross time point same lesion assessment as well as cross time point hottest lesion assessment.

There are 3 basic methods as follows:

1) Single most FDG-avid lesion: The most FDG-avid lesion at baseline that meets previously stated minimum requirements is defined on all time points. Relative change in this single lesion is calculated at each follow-up time point compared to baseline as follows:

$$\frac{SUV_{FU}(TL_{BL}) - SUV_{BL}(TL_{BL})}{SUV_{BL}(TL_{BL})}$$

Where: BL = Baseline scan
FU = Follow-Up scan
TL_{BL} = Target Lesion with greatest SUV at baseline

2) Single most FDG-avid lesion at each time-point: The most FDG-avid single lesion meeting minimum requirements is selected at baseline as well as each time point. The follow-up lesion is not necessarily the same lesion as the baseline lesion or other follow-up time points. The relative difference between the baseline target lesion (TL_{BL}) and the follow-up target lesion (TL_{FU}) is calculated as follows where the target lesions are not necessarily the same:

$$\frac{SUV_{FU}(TL_{FU}) - SUV_{BL}(TL_{BL})}{SUV_{BL}(TL_{BL})}$$

Where: TL_{FU} = Target Lesion with greatest SUV at follow-up

The workflow for the on-study evaluations is based on determining the most FDG-avid tumor lesion on each individual study independent of the baseline or any previous studies and performing the analysis and interpretation of the

most FDG-avid single lesion; thereafter finding the non-target lesions (lesions other than the most FDG-avid lesion) and performing the analysis and interpretation on those that are pertinent, if any; and finally performing the summary statistical interpretation on the per subject basis (as opposed to the per lesion basis).

3) Summed target lesions: Up to five most FDG-avid lesions are defined on the baseline examination (with no more than two per organ and all lesions meeting the defined metabolic threshold). The same target lesions are defined at each follow-up time point. For each time-point the sum of all target lesions is calculated. The change in the summed target lesions is calculated at each follow-up time point relative to baseline as follows:

$$\frac{\text{SUM } [SUV_{FU}(TL_i)] - \text{SUM } [SUV_{BL}(TL_i)]}{\text{SUM } [SUV_{bl}(TL_i)]}$$

Where TL_i = from 1 to 5 target lesions

The workflow for the on-study evaluations begins with finding the same lesions that were chosen as the target lesions on the baseline examination and performing the analysis and interpretation on each of them; thereafter finding the non-target lesions from the baseline examination and performing the analysis and interpretation on each of them; and thereafter finding any new lesions that meet the minimum threshold requirements and performing the analysis and interpretation on each of them; and finally performing the summary statistical interpretation on the per subject basis (as opposed to the per lesion basis).

The preceding workflow is contrasted with the workflow in the paradigm that depends on using the five most FDG-avid lesions as defined on each examination independently from one another (with no more than two per organ and all lesions meeting the defined minimum threshold), the workflow for the on-study evaluations begins with defining the five most FDG-avid lesions as previously defined without regard to the lesions chosen at baseline or any preceding studies and performing the analysis and interpretation of those five lesions; thereafter finding any pertinent non-target lesions (lesions other than the five most FDG-avid lesions) and performing the analysis and interpretation on those that are pertinent, if any; and finally performing the summary statistical interpretation on the per subject basis (as opposed to the per lesion basis).

The details for response assessment within each of these paradigms are specified in the subsequent Section 10.3. The definition of “the target lesion” should be based on the preceding criteria that include SUV measurement, reproducibility, measurability, motion, etc. The use of the response assessment paradigms is categorized by performance level as:

Acceptable –

- Option 1: Single target lesion at baseline followed over all subsequent studies (i.e. generally the most FDG-avid single lesion but defined as the same lesion from time point to time point).

- Option 2: Single target lesion (generally the most FDG-avid single lesion but potentially a different lesion from time point to time point provided that the lesions were both present on both studies – i.e. not a new lesion on the subsequent study(ies)).

Whichever option is chosen as the primary metric for the specific clinical trial, it is strongly suggested that data derived by both methods would be archived to allow post-hoc analysis of the clinical trial data.

Target –

- Option 1: In addition to the acceptable performance, sum of the most FDG-avid five target lesions with no more than two per organ (potentially different lesions from time point to time point) with all lesions meeting the minimum threshold requirements.

- Option 2: Most FDG-avid five target lesions at baseline followed over all subsequent studies (i.e. defined as the same lesions from time point to time point). This option may have utility when lesion selection is performed in the context of RECIST 1.1 anatomic response assessment criteria.

Whichever option is chosen as the primary metric for the specific clinical trial, it is strongly suggested that data derived by both methods would be archived to allow post-hoc analysis of the clinical trial data.

Ideal (exploratory) -
In addition to the acceptable and target (either Option 1 or Option 2) level of performance one would also determine the TLG activity across lesions included in the paradigm's dataset meeting the PERCIST minimum threshold (either only the five target lesions or all lesions, to be specified in the protocol). The use of TLG activity has not yet been validated across multiple tumor types in a multi-institutional setting. Hence, while this level of performance may be categorized as ideal, it is at this point in time exploratory in nature.

There may be alternative trial designs for specific clinical trial endpoints (e.g., targeting specific lesions based on local-regional therapies or correlation with biopsy).

10.2.2.2. Definition and Management of "New Lesions"

A new lesion is defined as either 1) an anatomic area that had no evidence of disease at baseline by FDG activity but with FDG activity on the follow up study AND a confirmatory anatomic lesion that is not related to a false positive cause (e.g. infection, treatment effect) or 2) an anatomic area that had no evidence

of disease at baseline by FDG activity but with FDG activity on follow up study but without a confirmatory anatomic lesion that is not related to a false positive cause (e.g. infection, treatment effect) that is confirmed as persistent at one-month follow up (by FDG and/or CT and/or biopsy). In the case of the latter definition, the dating of the new lesion should be the time of first appearance that met the previously defined minimum FDG-activity threshold. Some tumors might be anatomically new lesions without FDG activity. Non-FDG avid lesions should be assessed by RECIST 1.1 criteria [45]. For non-target lesions, please see Section 10.2.1.2.

10.3. Required Characteristics of Resulting Data – Summary Output Data (Response Assessment)

Objective response
Description of response should preserve the intrinsically continuous and quantitative nature of PET SUV. Determination if a response has occurred at all (i.e. if the quantitative alteration is greater than expected due to intrinsic biological variability and measurement error) is critical. It may also be convenient to further classify or categorize response (e.g. CMR, PMR, SMD, PD). Quantitative response metrics should be determined with consideration of multiple factors including, but not limited to, the purpose of the trial, the precise timing of the PET/CT scans within the imaging and treatment schedule (including the allowable window around each time point), the tumor type, the treatment paradigm employed, and the type(s) of decision(s) that will be based on the response assessment.

In particular, the choice of absolute or relative threshold for determining response category may depend on the context (e.g. % change may depend on tumor type and treatment). In addition, the utility and purpose of the response assessment will impact the appropriate threshold. For example, a larger threshold (e.g. => 30%) may be appropriate for predicting therapeutic efficacy and/or clinical evaluation of an individual patient, while a lower threshold (e.g. <=15%) may be appropriate for determining statistically significant change in a population of patients. Typically a larger change at the end of effective therapy is expected while smaller changes early after initiation of treatment may be indicative of response.

There are a number of proposed schemas (EORTC, PERCIST) available to guide the categorization of quantitative response metrics (as derived by methods described previously in Section 10 of this document), which are otherwise a continuous variable. Should the proposed schema include confirmatory imaging studies, the type and timing of such confirmatory imaging should be specified in the protocol.

The proposed response assessment schema references two comparator imaging timepoint scans: baseline scan and “best response” scan. The baseline scan timepoint is defined as the scan timepoint performed prior to initiation of the focused intervention under investigation. Thus, often the baseline scan is done prior to any therapy. However, when there has been prior therapy or there is a change in therapy, sufficient time should elapse following the prior therapy to ensure that the patient is in a stable state at the time of the baseline scan.

The best response scan timepoint is defined as the scan timepoint at which the lowest level of disease (or maximal response to the therapeutic intervention) is identified. The best response timepoint may be the same as the baseline timepoint if there is no interval (on-study) timepoint

that shows improvement. If progressive disease is determined using comparison to a nadir scan, then a follow-up confirmatory PET/CT scan is suggested. There is limited literature on progression and the zuse of comparisons to nadir, partially due to the small number of imaging time points.

Although RECIST criteria uses comparison to the best response or nadir of tumor size response, it is not clear that this approach should be used in assessing response using metabolic imaging. In some cases it may be appropriate, but at this time it is not clear that the concept of change compared to nadir response should be used with FDG imaging. The current recommendation is that comparison should be done compared to the baseline scan, which is obtained prior to any therapy, or to a baseline scan that is done once any acute response to prior therapy has resolved.

In some cases, particularly relatively early after start of therapy, FDG uptake in tumor can increase without reflecting true disease progression. This has been termed "pseudo-progression" [46, 47, 48, 49]. This only occurs in some settings, but must be considered in data interpretation in the design of a new clinical trial.

For assessment of a responder (CMR or PMR), comparison is made to the baseline timepoint. For assessment of progression (PMD), comparison can be made to either the baseline timepoint or the nadir timepoint. See section below on PMD for further discussion. If the nadir timepoint is used as the comparator for PMD and time to progression is being evaluated as a reportable value, then time zero should be defined as the time of the baseline timepoint. This calculation would then capture the time interval between initiation of focused intervention and time of progression.

One potential categorization schema is presented for consideration in this document (PERCIST). This schema also does capture the essence of the EORTC criteria.

Objective response reporting should be provided based on the following performance thresholds:

Acceptable:
The categorization schema used for a particular clinical trial should be clearly outlined in the clinical trial protocol prior to activation and data analysis. The rationale for the categorization schema used should be provided in the clinical trial design (which may be accomplished by reference to a societal standard or a publication in the peer-reviewed literature). Whichever categorization schema is used, the continuous un-categorized quantitative data as derived by methods described previously in Section 10 of this document should be retained and made available for post hoc analysis. Furthermore in cases of disease progression and/or response, data should be retained and made available regarding the quantitative and qualitative behavior of target, non-target, and new lesions including both PET and concomitant / follow-up CT-derived information.

Target and Ideal:
While total lesion glycolysis and tumor burden may provide additional information, there are insufficient data at this time to suggest the ideal method for assessing response.

An example categorization schema follows.

Progressive Metabolic Disease (PMD):

In a clinical trial that includes only a pre-intervention scan and a post-intervention scan, PMD is defined as significant increase in tumor uptake compared to baseline. Note that, particularly when imaging is done relatively early after treatment, increased uptake may indicate a good response (pseudo-progression).

In a clinical trial that includes multiple post-intervention scans (perhaps in trials with longer term follow up after completion of therapy) it is useful to compare tumor uptake to "best response" uptake values. In this case, PMD is defined as a significant increase in tumor uptake compared to "best response". It is acknowledged that progression from the baseline is a very conservative approach that may undercall the date of PMD. If the best response timepoint is prospectively defined as the comparator for PMD assessment in a protocol, then it is strongly suggested that a confirmatory follow-up time point be performed at least when progression is defined ONLY in terms of a rise in SUV (and not new lesions).

Progressive disease can be assigned based on progression of target lesions, identification of one or more new lesions or unequivocal progression of non-target lesions as further defined:

1) Target Lesion Assessment: It is proposed in PERCIST for the single most FDG-avid lesion at each time point (not necessarily the same lesion) that at least a 30% increase in FDG uptake, with ≥1.0 increase in SUV unit (or ≥0.8 increase in tumor SUL peak) be used as the threshold for PMD, given assurance of technical quality of scan. If more than one target lesion option is chosen, the sum of all target lesions (up to 5) at baseline and follow-up should be calculated and then this increase will be calculated as sum change of all qualifying target lesions identified, not based on any one of the target lesions; and/or

2) Non-target Lesion Assessment: Unequivocal progression of FDG-avid non-target lesion(s). There is currently no literature-based threshold defined to qualify the unequivocal requirement. Intuitively, the level of increase should probably be larger than that required for target lesion PMD to avoid overweighting of non-target assessment in PMD categorization. If PMD is based on non-target lesion assessment ONLY or primarily, then progression should be verified by confirmatory contemporaneous and/or follow-up imaging (which should be performed within 1 month) and/or biopsy unless PMD also is clearly associated with progressive disease by RECIST1.1; and/or

3) New Lesion Assessment: One or more new FDG-avid lesion(s) that are typical of cancer and not related to treatment effect, infection or inflammation; this typification may also require confirmatory studies in some circumstances. (See Section 10.2.2.2).

PMD should be reported to include percentage change in SUV units, (including, time after treatment, in weeks) and whether new lesion(s) are present/absent and their number. For example, rather than merely reporting PMD, the categorization should be specified to state that the SUV has increased by some value (e.g., +35%) as measured at some specific time point (e.g., week four) and the number if new lesions present at this time point if any (e.g., "in addition there are five new lesions). Because SUV is continuous variable, dividing response criteria into limited number of somewhat arbitrary response categories may result in loss of data. For this reason, PERCIST preserves percentage changes in SUV units in each reported category. Because rapidity with which the scan normalizes may be important (faster appears to be better), PERCIST asks for time from start of treatment as part of reporting. For example, a CMR with a change in SUV of -

90%, at one week, is probably superior to a CMR with a change in SUV of -90%, at ten weeks; especially if the latter subject was previously evaluated as SMD with a percentage change of SUV of -20% at the one-week post treatment evaluation.

As analysis of TLG volume is being proposed as an exploratory endpoint, this metric should not be used in isolation to determine PMD at this time. However, the data should be made available as previously stated (see Section 10.2.2.1).

Complete Metabolic Response (CMR):

1) Complete resolution of FDG uptake within measurable target lesion(s) so that the uptake is less than or indistinguishable from blood-pool levels (When liver activity is available for evaluation, this implies that the lesion uptake would be less than mean liver activity).
2) Disappearance of all other (i.e. non-target lesions) lesions to background blood pool levels.
3) Percentage change in FDG uptake should be recorded from the measurable region, as well as the time in weeks after treatment was begun. For example, in addition to reporting the CMR the report should also include the percentage change in SUV (e.g. -90%) and the time at which the evaluation is being made (e.g. four weeks). If there is both anatomic and functional complete response, there is no anatomic lesion to target for SUV measurement. Hence, a change in the SUV of the lesion is not possible to measure, especially if there is only one target lesion. Recording the background activity at the site of the previous lesion (provided there is no obvious artifact in the anatomic region) or the liver or blood background could be explored.
4) No new FDG–avid lesions in pattern typical of cancer.
5) If progression is noted by RECIST (anatomic measurement), but not by metabolic activity, verify with follow-up imaging.
6) There may be "faint" activity in certain lesions that is greater than immediate background but that is less than or indistinguishable from blood-pool levels. The presence of such lesions and the absolute SUV measurement should be noted; however, their presence should not dissuade classification as CMR provided those lesions meet the aforementioned criteria.

Partial Metabolic Response (PMR):

1) Reduction of minimum of 30% in target measurable tumor FDG uptake.
2) Absolute drop in SUV must be at least 1.0 (the absolute drop in SUL must be at least 0.8 SUL units), as well. Measurement is commonly in same lesion(s) as baseline but can be (an)other lesion(s) if the lesion(s) was previously present and is currently the most active lesion after treatment (see Section 10.2.2.1). ROI/VOI does not have to be in precisely same area as the baseline scan, though typically it is.
3) No increase equal to or greater than 30% in FDG uptake (must be at least 1.0 SUV or 0.8 SUL units, as well) or size of target lesion(s) (i.e. no PD by RECIST 1.1 or IWC) (if PD anatomically, must verify with follow-up). Reduction in extent of tumor FDG uptake is not requirement for PMR. Percentage change in SUL should be recorded, as well as the time in weeks after treatment was begun. For example the categorization as PMR should be further qualified by including the percentage decrease in SUV units (e.g. -40%) and the number of weeks after treatment initiation at which the observation is made (e.g. three weeks).
4) No new lesions.

Stable Metabolic Disease (SMD):

1) Not CMR, PMR, or PMD.

2) SUVpeak in metabolic target lesion(s) should be recorded, change in SUVpeak of the target relative to the baseline , as well as the time from start of most recent therapy, in weeks. As has previously been suggested the categorization as SMD should be accompanied by the percentage change in SUV units (e.g. -15%) and the number of weeks after treatment initiation at which the measurement is made (e.g. seven weeks).

Overall Best Response in a given subject (summation of time point determinations using the categorization schema above including target and non-target lesions; new lesion; etc.):

Best time-point response (e.g. CMR, PMR, SMD, PMD) that is noted during the time period defined as the time from treatment start to:

1) CMR, or
2) Disease progression / recurrence, or
3) Termination of the subject from the clinical trial.

Duration of Best Response in a given subject (summation of time point determinations using the categorization schema above including target and non-target lesions; new lesions; etc.):

1) Measured from the date Best Subject Response criteria are first met to date disease progression / recurrent disease is first noted or the date that the subject has completed the trial follow-up period (with some indication that the Best Response category (e.g. CMR, PMR/SMD) may still be ongoing). Note, CMR by RECIST 1.1 is not required. However, the criteria for CMR for the specific trial should be specified in the clinical trial documentation. Progression from PMR to PMD is suggested (i.e. the transition from PMR to SMD may be insufficient) to end the "Duration of Best Subject Response" for subjects with PMR as the transition from PMR to SMD may not be clinically relevant and/or statistically robust.
2) **Duration of Overall Response in a given subject**: from date CMR and/or PMR criteria are first met (whichever status came first); to date PMD is first noted or the date that the subject has completed the trial follow-up period (with some indication that the best overall response category may still be ongoing). Progression from PMR to PMD is suggested (i.e. the transition from PMR to SMD may be insufficient) to end the "Duration of Best Subject Response" for subjects with PMR as the transition from PMR to SMD may not be clinically relevant and/or statistically robust.
3) **Time to Progression:** from date of treatment start to date PMD is first noted by PET/CT.
4) **Duration of SMD**: In subjects that do not achieve an observed CMR or PMR, the Duration of SMD is defined as the time from initiation of therapy to the time of PMD.
5) **Progression Free Survival**: defined as the time from the initiation of therapy to the time of PMD or death. Progression Free Cancer-specific Survival is measured from the time of therapy initiation to the time of PMD or death due to cancer.
6) **Note:** If PMD must be confirmed on a follow up scan for any of these measures of duration, PMD would be timed to the date when PMD was FIRST noted by PET/CT criteria, not the date of confirmation.

10.4. Platform-specific instructions {*This section intentionally omitted*.}

10.5. Reader Training

Reader training should be specified in the clinical trial documentation for the specific clinical trial or reference may be made to generic reader training documents when appropriate.

10.6. Archival Requirements (See 11.7.)

10.7. Quality Control (See 12.7.)

11. Archival and Distribution of Data

11.1. Central Management of Imaging Data

Two sources [3, 8] mention use of DICOM formatted data. One source [3] indicates that data should be stored in DICOM format Part 10: Media Storage and File Format for Media Interchange. DICOM format should meet the Conformance Statement written by manufacturer of the PET/CT system [3].

Acceptable: Data should be stored and transmitted in compliance with pertinent DICOM standards (which for CD and DVD storage and transmission is DICOM format Part 10: Media Storage and File Format for Media Interchange). When data are transmitted using ftp or other Internet-based systems, the archival and transfer method used must allow transmission of all data necessary for qualitative and quantitative assessments without alteration of the data from the acquisition state. All data transfer should be secure and HIPAA-compliant. When a central archival and review facility is used in a clinical trial, the individual trial design should explicitly state what types of data (e.g. raw data, reconstructed data, post-processed data, etc.) are to be transmitted to the central facility in addition to being archived at the participating site.

11.2. De-identification / Anonymization Schema(s) to Be Used

Two sources [3, 8] indicate that DICOM image data need to be de-identified/anonymized. The header of the DICOM formatted images may contain information that identifies the patient and these tags should be scrubbed or these tags may be replaced by information about study ID, randomization or case IDs as indicated by the image core lab. De-identification must be performed prior to transmittal of the data from the local site to the image core lab. Both sources indicate use of (s)FTP as means of transmittal. One source [3] suggests storing de-identified DICOM formatted images on media (CD, DVD) and sending it by regular mail.

Acceptable: Data de-identification / anonymization is performed on a third-party or PACS workstation in a manner that is HIPAA-compliant and compliant with the directions of the clinical trial. However, all data necessary to perform qualitative and quantitative assessments must remain available and unaltered. Hence, removal of PHI should not affect the underlying imaging data. Specifically all data necessary for reconstruction, post-processing, interpretation, and analysis should not be affected by the removal of PHI during the de-identification process. And any algorithms used for de-identification should not remove prerequisite imaging data when PHI is removed. There needs to be a mechanism to perform quality control to ensure that the de-identified / anonymized imaging data correctly correspond to a specific subject ID.

Target / Ideal: In addition to the acceptable performance level, data de-identification / anonymization is performed on the image acquisition platform in a manner that is HIPAA-compliant and compliant with the directions of the clinical trial. There is no admixture of PHI and imaging data within the same DICOM fields. There should be no PHI in private fields (i.e. DICOM tags). There should be no imaging data necessary for qualitative or quantitative assessments in private fields (i.e. DICOM tags).

11.3. Primary Source Imaging Data

Acceptable: All FDG-PET/CT studies used within the context of the clinical trial should be archived as primary source data and should be subjected to the quality assurance mechanism for imaging obtained within the context of the clinical trial. Archival of raw projection data is optional. If raw projection data are of interest for a particular trial, the trial protocol should state explicitly the standards for the format and storage (including the duration of storage) of such data. All archives and archival processes should be secure and should include disaster recovery.

Target / Ideal: In addition to the acceptable level of performance, archival of raw projection data is also mandated in a secure and redundant manner for a duration the same as for all other archived trial data.

11.4. Reconstructed Imaging Data

Acceptable: Archival of reconstructed image data either by DICOM format Part 10-compatible media storage or local PACS / server-based storage by both the sites and the central review entity (if any). Archival of raw projection data is optional. If raw projection data are of interest for a particular trial, the trial protocol should state explicitly the standards for the format and storage (including the duration of storage) of such data. All archives and archival processes should be secure and should include disaster recovery.

Target / Ideal: In addition to the acceptable level of performance, archival of raw projection data is also mandated in a secure and redundant manner for a duration the same as for all other archived trial data.

11.5. Post-Processed Image Data

Acceptable: If post-processed image data is included in the clinical trial imaging protocol or is used during the analysis and interpretation steps whether specified in the trial protocol or not, such post-processed image data should be archived at the time and by the site at which the post-processing is performed, inclusive of all data that was used in the post-processing.

11.6. Analysis Results

Acceptable: Archival of the analysis is performed at the time and by the site at which the analysis is performed by use of a clinical trial-specific case report form that references the specific slices and lesions and provides all pertinent qualitative and quantitative data as required by the clinical trial protocol. DICOM secondary image capture may be optionally included for clarification.

Target: In addition to the acceptable level of performance, archival of the analysis is performed at the time and by the site at which the analysis is performed by use of annotations and/or mark-ups on the reconstructed (or post-processed) image data and saved as a new series so that the original reconstructed (or post-processed) image data are retained without alteration. Theses annotations and/or mark-ups may be archived either as a "screen save" or DICOM secondary image capture.

Ideal: As per Target, except the ROI / VOI data are captured as true primary data in DICOM format rather than as a representation of the ROI / VOI data captured as an image.

11.7. Interpretation Results

Acceptable: All site interpretation results (see Section 10) should be archived at the time and at the site that such data output is generated. When a central facility is included in the trial design, the site interpretation results and the central facility interpretation results should be archived at the central facility. These results include, but are not limited to, the interpretation and analysis data output as described in detail within Sections 9 and 10 of this UPICT Oncologic FDG-PET/CT protocol pertinent to the clinical trial design. Merely archiving the summary statistics at the subject level over all time points is considered insufficient for QA and reproducibility assurance. The duration of archive for the imaging data should be the same as for all other trial-related data unless otherwise stipulated by the sponsor and/or regulatory oversight agencies.

12. Quality Control

12.1. QC Associated with the Site

12.1.1. Quality Control Procedures

The Imaging QC section of the clinical trial protocol should specify how site compliance should be verified and documented. There should be specific site report forms and checklists to facilitate the verification and documentation of QC.

If exceptions to any of the performance standards stated below occur and cannot be remediated on site, the site should promptly communicate the issue to the appropriate internal overseer / coordinating center / core lab for advice as to how the irregularity should be managed; if possible this communication should occur prior to acquisition of any subject data.

All Target performance specifications are in addition to those stated for the Acceptable level of performance. Similarly, all Ideal performance specifications are in addition to those stated for both the Target and Acceptable levels of performance.

All auxiliary equipment (e.g. clocks, scales, stadiometer, glucometer, and dose calibrators) are calibrated and/or synchronized and/or periodically monitored and documented as part of an ongoing QC program as follows:

12.1.1.1. Clock Calibration and Synchronization:

Acceptable: Checks for internal consistency daily and after service events. Synchronization of all clocks used in the conduct of the FDG-PET/CT study should be performed monthly or as needed based on consistency checks. Dose calibrator and scanner computer clocks and all clocks used in the conduct of the imaging study are synchronized within +/- 60 seconds.

Target: Checked weekly against an external reference standard (e.g. NTP or equivalent appropriate standard at the site of acquisition).

Ideal: Dose calibrator and scanner computers are synchronized daily through a vendor-supported automated process against the reference standard and therefore within +/- 5 seconds of reference standard.

12.1.1.2. Scales and Stadiometer Calibration and Performance:

Acceptable: Verified at the time of installation/commissioning and checked on a regular basis (no less frequently than annually) by assigned institutional staff.

Ideal: Required data is transferred directly from measurement device into scanner by electronic, HIS/RIS, or other means bypassing operator entry but still requiring operator verification.

12.1.1.3. Glucometer Calibration:

Acceptable: Glucose measurements should be made using a CLIA approved, CLIA cleared, or equivalent (outside the US) glucose measurement technique.

Ideal: Required data is transferred directly from measurement device into scanner by electronic, HIS/RIS, or other means bypassing operator entry but still requiring operator verification.

12.1.1.4. Dose Calibrator(s) QC:

Acceptable: All calibration tests are performed per the manufacturer's directions and as defined by the applicable regional and national regulatory bodies using acceptable reference standards (e.g. NIST). The most recent manufacturer-specific F18 gain settings are used during these calibration tests. Accuracy, linearity, and geometry tests should be performed at installation and after service events. Linearity testing should be performed at least quarterly. Accuracy testing should be performed at least annually using the appropriate reference standard. Daily constancy should be measured with a long-lived isotope in the range of 500-650 keV and net measured activity should be within +/- 5% of expected value. Manufacturer-recommended QC should be performed on dose calibrators that are part of an automated injection system. Cross calibration between manual dose calibrators that are used for scanner QC and/or manual injections and automated injection systems should be confirmed to be within 5%. Careful attention should be made to ensure

consistent injection technique including tubing length and diameter. It should also be confirmed that all of the activity is injected into patients following the designated flush.

Target: QC procedures should incorporate the use of traceable NIST (or equivalent) Ge68-calibration source to perform accuracy test at least annually to verify the F-18 calibration with deviation <+/-3%. Linearity testing should be performed quarterly using decay or attenuating sleeve method. Dose calibrators should be adjusted

Ideal: An NIST-traceable (Ge68 or other equivalent source) F18-simulation source is used to calibrate the dose calibrator calibration setting for F18 to match the reading to the actual activity of the NIST source. Required data is transferred directly from measurement device into scanner by electronic, HIS/RIS, or other means bypassing operator entry but still requiring operator verification.

12.1.1.5. CT component of PET/CT scanner

Acceptable: CT scanners require rigorous acceptance testing and routine QC to ensure appropriate image quality and radiation exposure. As these devices administer radiation, there are additional regulatory requirements at the national and/or state level. In addition, specific QC procedures should be performed according vendor recommendations. Examples or vendor-recommended CT QC procedures are shown. As an example of general procedures that should be formed on all scanners, the NCIE CQIE guidelines of CT QC are listed as follows.

Daily QC: At a minimum, daily QC should be performed prior scanning and include air calibrations, measurements of water CT numbers and standard deviations, and check for absence of artifacts.

Annual QC: The following tests should be performed at installation, after tube replacement, and annually:

- Scout Prescription & Alignment Light Accuracy
- Imaged Slice Thickness
- (slice sensitivity profile, SSP)
- Table Travel/Slice Positioning Accuracy
- Radiation Beam Width
- High-Contrast (Spatial) Resolution
- Low-Contrast Sensitivity and Resolution
- Image Uniformity & Noise
- CT Number Accuracy
- Artifact Evaluation
- Dosimetry/CTDI

Ideal: The results of QC testing should be exported in a file format that is readily accessible along acceptable ranges of performance.

12.1.1.6. PET Scanner or PET component of PET/CT scanner
(General QC Procedures including Calibration)

Acceptable: Scanner is cross-calibrated with same dose calibrator used to assay patient injections. The cross calibration should be reviewed/performed at least every 3 months, after scanner upgrades , after new setups, and after modifications to the dose calibrator (per ACRIN CQIE guidelines) [28].

The same scanner with the same acquisition/reconstruction protocol, software and settings should be used for each subject study. Only if the primary scanner is unavailable, a scanner demonstrated as having equivalent output (as predefined by the clinical trial site qualification and QC documentation and supported by accepted international standards) and qualified through the protocol's site qualification process may be used (ideally the second scanner should be of the same make, model, and software version as the primary scanner). The same scanner acquisition and reconstruction parameters should be used for QC as are being used for subject image acquisition (except for scan duration which may be extended for QC purposes).

Scanner calibration factors (as defined by each manufacturer specific to each scanner model) should be recorded and monitored. Variances of more than 3-5% are potentially due to mis-calibration and therefore should result in verification of correct calibration and/or recalibration as necessary.

At a minimum, phantom calibration should be performed annually using acceptable standards as enumerated below. The same method should be used by each site for the duration of the trial (not necessary for every site to use the same method).

A) ACRIN / EANM criteria for uniform cylinder [3, 50]
Overall Mean Bkgd. SUV = 1.0 ± 0.1

B) Modified ACR phantom criteria (note the modification of SUV Bkgd criterion)
1. Mean Bkgd SUV: 0.9 – 1.1
2. 25 mm cylinder: > 1.8 – < 2.8
3. 16 mm / 25 mm ratio: > 0.7

C) SNM CTN criteria
1. SUV = 1.0 ± 0.1 as assessed in the standard uniform portion of the standard CTN oncology phantom.
2. Visualization of all simulated lesions =>10mm.
3. SUVmax of simulated lesions 15mm or 20mm >= 2.2.

D) NCI CQIE
1. Volume-averaged SUV in phantom between 0.90 and 1.10
2. Axial variation in phantom < 10%

3. Dynamic studies: Volume-averaged SUV of each time frame varies by < 10% over the course of the 25-minute acquisition.

Manufacturer specific Image registration calibration between the PET and CT scanner should be performed at installation and after service events that involve moving either device. The image registration should be evaluated annually or after any suspicion of mis-registration. Registration calibration should be performed after any confirmed mis-registration that exceeds the manufacturer's specified tolerance.

Target: Scanner calibration, uniformity and recovery coefficient versus sphere or cylinder diameter should be assessed quarterly or after any major service or upgrades that may affect quantitative accuracy.

Ideal: Each site shall perform and document the full range of the QC tests listed below (as specified by the **Ideal** performance characteristics) using automated, standardized methods and phantoms (i.e., those listed above) to document compliance. This should be part of site qualification and then should be repeated periodically, at least annually and after any major service and after any scanner recalibration related to software upgrades. Vendors should implement daily quality control reports that can be exported and submitted along with patient studies for clinical trials.

SUV measurements for a standardized phantom should have an overall mean SUV = 1.0 ± 0.05. ROIs (approximately 4 cm or greater but not including portions subject to partial volume effects) appropriate to the use instructions for the particular phantom employed.

Cross calibration with dose calibrator is accomplished with paired NIST-traceable sources for the dose calibrator and PET scanner. This calibration is checked weekly.

Image registration between PET and CT images should be evaluated periodically including the effect of patient weight and bed deflection.

12.1.1.7. Syringes and tubing used during QC processes:

Acceptable: Syringes and injection tubing are assayed pre- and post-injection and pertinent information (i.e. time of measurement and amount of residual activity) is recorded routinely if applicable to the specific scanner QC routine and capabilities. The injection technique should be standardized by ensuring that the same specification of syringes and tubing are used.

12.1.1.8. Normalization:

Acceptable: Normalization of detector response should be performed according to vendor recommendations at least every 3 months, after relevant service events, after appearance of software/hardware upgrades, and

appearance of artifacts in uniformity check. Vendor-specific quality daily control checks should be performed and confirmed to be acceptable.

Target: Documentation of the normalization and results should be provided in a readily accessible format.

Ideal: For some systems, more frequent normalization may be preferred (e.g. monthly) provided that this is done in an automated manner with minimal risk of human error.

12.1.1.9. Uniformity:

Acceptable: In addition during the normalization and calibration methods outlined above, transverse and axial uniformity should be assessed with a uniform phantom using a water phantom with F18 at least every 3 months, after new scanner calibrations, and after software upgrades. Qualitative review should be performed (i.e. by visual inspection) to ensure that there are no artifactual variations within or between axial slices.

Uniformity should be assessed with a uniform cylinder with an F-18 compound in water. For uniformity tests the cylinder can also use Ge-68/Ga-68 in epoxy as a sealed solid source, but only if the uniformity has been verified by other means. The ROI employed should conform with the use instructions for the particular phantom employed. Phantom quantitative measurements with overall mean SUV = 1.0 ± 0.10 should be made with an ROI (approximately 3 cm or greater but not including portions subject to partial volume effects) appropriate to the use instructions for the particular phantom employed.

By ACRIN/EANM/SNM criteria axial slice uniformity does not vary more than 10% from one end of the axial FOV to the other.

By SNM CTN criteria, phantom sections of uniformity do not vary more than 10% from one another.

Target: The overall mean SUV = 1.0 ± 0.05 should be made with an ROI (approximately 3 cm or greater but not including portions subject to partial volume effects) appropriate to the use instructions for the particular phantom employed.

Ideal: Daily uniformity measurements are performed and recorded in an accessible manner that can be exported and distributed with individual patient studies.

12.1.1.10. Image Quality:

Acceptable: A standardized image quality phantom scan should be performed at least annually to check hot and cold spot image quality per the ACRIN CQIE guidelines [28]. Additional review of resolution and noise should be performed

according to specific trial guidelines and as stated below. Currently there is no consensus phantom that should be used. CT and PET co-registration should meet the manufacturers' recommendations at scanner acceptance and after any major service events that involve moving scanner gantries.

For individual patient studies, qualitative assessment should be performed to evaluate co-registration, noise, resolution, and other aspects of image quality (see 9.5.1.1). See sections below for specifics aspects of (resolution and noise).

Target/Ideal: Minimum standards for image quality should be defined based on the requirements of specific trials. Ideally co-registration should be inspected visually with a weight load to evaluate bed deflection due to patient weight.

12.1.1.11. Resolution / SUV Recovery:

Acceptable: At a minimum annually, each site shall perform and document a qualitative resolution QC test by using the manufacturer's settings and demonstrating resolution of normal gross anatomic features within clinical images of the brain, heart, and abdomen (e.g. the images should not appear "too smooth").

Per SNM criteria and using the CTN PET Oncology Phantom [9] (and based on the use of the site's standard clinical acquisition and reconstruction protocols), all lesions 10mm or greater should be visually detectable for those sites that have access to this phantom. For sites without access to this phantom an equivalent quantitative test should be performed.

The ACR criteria for resolution [28] (based on the use of the site's standard clinical acquisition and reconstruction protocols) are:

- The lower portion of the cylinder contains six sets of acrylic rods arranged in a pie-shaped pattern with the following diameters: 4.8, 6.4, 7.9, 9.5, 11.1, and 12.7 mm.
- At this target level, the 9.5, 11.1, and 12.7 mm diameter rods must be visible.
- By ACR criteria, resolution should be achieved as measured by a 25 mm cylinder is >1.8 and <2.8 3 or by a 16/25 mm cylinder ratio: >0.7 Ref ACR PET phantom test guidelines (revised 2/22/10).

The EANM guidelines [3, 9] provide harmonizing performance criteria for SUVmax and mean recovery as function of sphere size (NEMA NU 2 2007 IQ phantom) and thereby should ensure comparable quantitative scanner performance between sites.

For information on the SNMMI/CTN phantom please see the SNMMI/CTN website [51]. Using the CTN PET Oncology Phantom, the scanner resolution is

assessed by ensuring that all lesions =>10mm are visually detectable and that lesion SUVmax values are within an acceptable range.

Target: Scanner reconstruction protocols are adjusted to provide at least appropriate resolution properties as defined for the specific trial (i.e. recovery coefficient versus sphere or cylinder diameter) for a standard test object (e.g. ACR cylinders or NEMA spheres or other similar phantoms) that contains specific "hot spot" objects [11].

Ideal: Vendors implement a reconstruction protocol that ensures pre-defined image recovery coefficient characteristics are met. This implementation has two components. The first component is that every site in a particular trial and preferably across all trials would use the same calibration methods / phantom as prescribed in an accepted standard (either the same methods and phantom or the same methods coupled with a defined set of phantoms that have equivalent performance characteristics.

The second component is that the vendors would provide or support the users to implement an acquisition / reconstruction protocol that produces the desired results and the vendors provide an automated image assessment tool to verify that the acquisition and reconstruction protocols produce the desired results.

12.1.1.12. Noise:

Acceptable: During routine testing, e.g. done as a regular QA or QC procedure or for qualification purposes, and when the site uses the trial-specific acquisition parameters (e.g. time per bed position, dose, reconstruction etc.), the noise in phantom images should be assessed qualitatively to be of consistent and acceptable quality.

Target: During routine testing, e.g. done as a regular QA or QC procedure or for qualification purposes, and when the site uses the trial-specific acquisition parameters (e.g. time per bed position, dose, reconstruction etc.), the noise in phantom images should be measured by reporting the mean, standard deviation (SD), and COV of voxel values within a volume of interest (VOI) as described in section 7.2.

Images are reconstructed with a voxel size of 3-4 mm all three dimensions, but not necessarily isotropic.

Ideal: During routine testing, e.g. done as a regular QA or QC procedure or for qualification purposes, and when the site uses the trial-specific acquisition parameters (e.g. time per bed position, dose, reconstruction etc.), the noise in phantom images should be measured by reporting the mean, standard deviation (SD), and COV of voxel values within a volume of interest (VOI) as described in section 7.2 [11].

12.1.2. Baseline Metrics Submitted Prior to Subject Accrual -- *See section 12.1.1.*

Acceptable: Representative human subject images consistent with the specifics of the clinical trial should be carefully examined to finalize site qualification. This may be accomplished by one of several strategies. For example, one strategy would be to require submission of patient studies performed prior to the trial and outside of the trial. A second potential strategy may be to require rigorous QC review of the first one or two accrued subjects in the context of the trial. A third potential strategy would be to include initial "human subjects imaging" on subjects not getting the targeted intervention but obtained purely for the purposes of site qualification for the study. A combination of these mechanisms might also be used. Whatever mechanism is used should be compliant with human subject protection regulations and the sites' IRB requirements.

12.1.3. Metrics Performed *and/or* Submitted Periodically During the Trial -- *See section 12.1.1.*

Acceptable / Target: The results of the QC procedures performed per Section 12.1.1. should be provided at least annually and should be available for any site audit. Should a new PET/CT system be installed, that equipment must be qualified for the trial if it is to be used in the trial. Any PET/CT system that undergoes a major upgrade (i.e., an upgrade that may affect the SUV determination) during the trial must be re-qualified prior to use in the trial.

Ideal: Variances in performance characteristics that remain within the range of normal but exceed a pre-specified threshold of percentage change should be documented and data should be aggregated for later analysis.

12.2. QC Associated with Imaging-related Substance Preparation and Administration

Acceptable: FDG must be obtained from a source that is approved by the geographically appropriate regulatory mechanism (e.g., in the USA an FDA-submitted NDA or ANDA). For geographic sites that lack such regulatory oversight, equivalency to the USA FDA NDA or ANDA standards is required.

12.3. QC Associated with Individual Subject Imaging (performed per subject or performed daily and therefore available for association with individual subject imaging)

12.3.1. Phantom Imaging and/or Calibration

Acceptable: None

Target: Daily phantom uniformity and calibration testing using Germanium cylindrical source or equivalent per manufacturers' specifications.

Ideal: Daily phantom uniformity, resolution, noise, and calibration testing using a F18 - fillable source* or a Germanium-68 cylindrical source or equivalent per manufacturers specifications.

*If an F-18 fillable phantom is used, there may be more human error associated with the procedure and hence use of a Germanium-68 cylindrical source is preferred.

12.3.2. Quality Control of the Subject Image and Image Data

Consolidated Statement – The integrity of DICOM image headers should be reviewed and confirmed for regulatory compliance (HIPAA), protocol compliance, and consistency with source data such as CRFs. In some cases, internal references such as the liver can be used for quality control to confirm acceptable ranges of SUVs [6].

Acceptable:

1. QC tests as described in sections 12.1.1 - 12.1.3 pertinent to the QC of the subject image data (i.e. visual qualitative inspection, alignment, motion artifact, noise, etc.)

2. DICOM header integrity and compliance with protocol and institutional / other policies (e.g. for multi-site trials HIPAA compliance), consistency with CRF data.

3. Internal QC control should be performed consistent with the performance standards expressed in Section 9.2.1.

4. Syringes and injection tubing are assayed pre- and post-injection and pertinent information (i.e. time of measurement and amount of residual activity) is recorded and is consistent with the data used for quantitative analysis.

Noise: When the site uses the trial-specific acquisition parameters (e.g., time per bed position, dose, reconstruction etc.), the noise in patient images should be assessed qualitatively to be of consistent and acceptable quality. I.e., the images should not appear too noisy' for trial-specific purposes.

Target (in addition to Acceptable):
Noise: When the site uses the trial-specific acquisition parameters (e.g. time per bed position, dose, reconstruction etc.), the noise in patient images should be measured by reporting the mean, SD, and COV within a VOI using methods as described in Section 7.2. The VOI should be positioned in the mid or lower region of the right liver.

Ideal (in addition to Acceptable and Target):
Noise: When the site uses the trial-specific acquisition parameters (e.g., time per bed position, dose, reconstruction etc.), the noise in patient images should be measured as described immediately above. The COV of the voxel values thus determined should be recorded and should be below 15%.

12.4. QC Associated with Image Reconstruction

Consolidated and Consensus Statement – Acceptable: CT images should be reviewed for potential artifacts such as beam hardening, metal objects, and motion. PET images should be compared to the CT images for proper image registration and potential attenuation correction artifacts [6]. PET images should be checked for other reconstruction artifacts

12.5. QC Associated with Image Post-processing

Acceptable: QC plan should be based on the type of post-processing that was performed (i.e., DICOM Header manipulation including, but not limited to de-identification tasks; post-processing that affects quantitation; and/or post-processing that affects visualization). The rigor of the QC process should be commensurate with the type of post-processing that was performed and the potential for unintended consequences associated with the post-processing performed. The QC process employed for post-processing tasks should be described in sufficient detail to allow "downstream" consumers of the trial data to have the necessary confidence in the imaging data for the purposes intended. The description of the QC process should be sufficiently detailed to allow non-trial personnel to perform validation checks of the QC process should they so desire.

12.6. QC Associated with Image Analysis

Acceptable: The imaging protocol should include a QC program for Image Analysis whether analysis is performed at a core facility, the acquisition sites, or both. Whatever program is stated should be followed and documented.

12.7. QC Associated with Interpretation

Acceptable: The imaging protocol should include a QC program for Image Interpretation whether interpretation is performed at a core facility, the acquisition sites, or both. Whatever program is stated should be followed and documented.

13. Imaging-associated Risks and Risk Management

13.1. Radiation Dose and Safety Considerations

The radiation dose of the PET/CT study results from radiation exposure from the injection of FDG and from the CT study [3, 12]. One source [3] indicates that CT scans can be performed as low dose CT to be used for attenuation correction purposes to minimize radiation dose. Two sources [3, 12] indicate that radiation dose from the CT scans should be estimated specific to the system and imaging protocol used [3] or by means of standard estimates. [5] These standard estimates can be utilized within the framework of local regulatory requirements for risk analysis [12], which will also depend on patient populations and life expectancy [12][5] and particular considerations to reduce radiation exposure should be given for pediatric applications [3]. There are several publications reporting radiation doses for FDG. A paper that summarizes both adult and pediatric doses is Alessio et al, 2009 [52]. For a typical administered dose of 370 MBq the estimated whole body radiation dose is 7 mSv. There is greater variability in the radiation doses from CT, which is very dependent on the exact protocol used (e.g. 1. CT for attenuation correction only, 2. CT with

improved anatomic localization, or 3. diagnostic CT). A recent study [53] suggests that the CT doses can range from 7 to 26 mSv. Many hardware and software improvements that have been developed for dose reduction in diagnostic CT studies are being used in PET/CT such as automated tube current modulation and iterative reconstruction. For pediatric studies, a common approach is to reduce kVp and tube current. Alessio et al. suggest that, with care it is feasible to decrease the CT doses to 3 to 6 mSv [52]. Particular consideration to reduce radiation exposure should be given for pediatric patients. One common approach in children is to administer approximately 5.3 MBq/Kg of FDG with a minimum dose of 37 MBq and a maximum dose of 370 MBq.

Acceptable / Target: The protocol and the informed consent form should contain language describing the estimated administered dose range and estimated whole body radiation exposure (expressed as effective dose in mSv) for the FDG to be administered. In addition both documents should provide comparator (equivalency) radiation examples. The estimates of radiation dose will be site and protocol-specific and based on factors such as the number and frequency of studies. Useful comparators are annual background radiation (~ 3 mSV/yr) and the allowable dose to radiation workers (50 mSv/yr).

Ideal: In addition to the above, each site should document the estimated radiation dose for each subject (whole body) inclusive of FDG and CT. The protocol should contain the estimated critical organ dose attributable to FDG based on the proposed administered dose.

13.2. Imaging Agent Dose and Safety Considerations

There is a potential small risk of allergic reactions, but there have been no reports of such reactions associated with intravenous administration of FDG.

Approximately 1 person in 1000 may have an allergic reaction from the iodinated contrast drugs. These reactions are temporary and treatable. Allergic reactions may include: mild itching or hives (small bumps on the skin), and shortness of breath and swelling of the throat or other parts of the body. The subject should be instructed to tell the technologist immediately if s/he experience any of these symptoms so s/he can be treated promptly.

The placement of intravenous catheters has the associated risk of making the patient temporarily uncomfortable and a small bruise may form. A slight bruise may form where the needle has been in a vessel. There is a slight risk of infection at the site, but sterile technique reduces this risk nearly completely. The patient may also experience claustrophobia from the imaging ring apparatus or discomfort from lying on the scanner table for 60-120 minutes.

Acceptable: The protocol and informed consent form should contain language stating that there have been no serious reported reactions to FDG. If iodinated contrast is used in the study, the protocol and informed consent should contain language outlining the risks associated with that contrast. The risks of intravenous access and the potential of extravasation of FDG and iodinated contrast should also be included in the protocol and informed consent document.

13.3. Imaging Hardware-specific Safety Considerations

Acceptable:
Per recommendations from the FDA, before beginning the first CT portion of the PET/CT scan, the operator should use history, physical examination, and CT scout views to determine if implanted or externally worn electronic medical devices are present and if so, their location relative to the programmed scan range.

For CT procedures in which the medical device is in or immediately adjacent to the programmed scan range, the operator should:

- Determine the device type;
- If practical, try to move external devices out of the scan range;
- Ask patients with neurostimulators to shut off the device temporarily while the scan is performed;
- Minimize x-ray exposure to the implanted or externally worn electronic medical device by:
    - Using the lowest possible x-ray tube current consistent with obtaining the required image quality; and
    - Making sure that the x-ray beam does not dwell over the device for more than a few seconds

After CT scanning directly over the implanted or externally worn electronic medical device:

- Have the patient turn the device back on if it had been turned off prior to scanning.
- Have the patient check the device for proper functioning, even if the device was turned off.
- Advise patients to contact their healthcare provider as soon as possible if they suspect their device is not functioning properly after a CT scan.

13.4. Management and Reporting of Adverse Events Associated with PET radiopharmaceutical or CT contrast agent

Acceptable: Adverse event (AE) tracking and reporting for FDG-PET/CT in the course of a clinical trial should be embedded in the general trial AE tracking and reporting mechanism. It is reasonable to limit the time frame for possible AE attribution to less than twenty-four (24) hours after administration.

13.5. Management and Reporting of Adverse Events Associated with Image Data Acquisition
Does not apply to this protocol.

APPENDIX A

ACRONYMS AND ABBREVIATONS

ACRIN:	American College of Radiology Imaging Network
AE:	Adverse Event
ANDA:	Abbreviated New Drug Application
CT:	X-ray Computed Tomography
CTDI:	CT Dose Index
DICOM:	Digital Imaging and Communications in Medicine
DLP:	Dose-Length-Product
EORTC:	European Organisation for Research and Treatment of Cancer
EU:	European Union
FDG:	Fluorodeoxyglucose
GHS:	Global Harmonization Summit
HIPAA:	Health Insurance Portability and Accountability Act
IRB:	Institutional Review Board
kVp:	Peak Kilo-voltage
mAs:	milliamp-seconds
MIP:	Maximum Intensity Projection
MTV:	Metabolic Tumor Volume
NDA:	New Drug Application
PET:	Positron Emission Tomography
PERCIST:	PET Response Criteria in Solid Tumors
PHI:	Protected Health Information
RECIST:	Response Evaluation Criteria in Solid Tumors
RSNA:	Radiological Society of North America
QA:	Quality Assurance
QC:	Quality Control
QIBA:	Quantitative Imaging Biomarkers Alliance
ROI:	Region-Of-Interest
TLG:	Total Lesion Glycolysis
UPICT:	Uniform Protocols for Imaging in Clinical Trials
VOI:	Volume-Of-Interest

APPENDIX B

REFERENCES

1. Young H, Baum R, Cremerius U, et al. Measurement of Clinical and Subclinical Tumour Response Using [18F]-Fluorodeoxyglucose and Positron Emission Tomography: Review and 1999 EORTC Recommendations. Eur J Cancer 1999, 35(13):1773-1782.

2. Shankar LK, Hoffman JM, Bacharach S, et al. Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials. J Nucl Med 2006;47:1059-66.

3. Boellaard R, O'Doherty MJ, Weber WA, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. Eur J Nucl Med Mol Imaging 2010;37:181-200.

4. ACRIN Protocol #6671: Utility of preoperative FDG-PET/CT scanning prior to primary chemoradiation therapy to detect retroperitoneal lymph node metastasis in patients with locoregionally advanced carcinoma of the cervix (IB2, IIA≥4 CM, IIB-IVA) or endometrium (grade 3 endometrioid endometrial carcinoma; serous papillary carcinoma, clear cell carcinoma, or carcinosarcoma (any grade); and grade 1 or 2 endometriod endometrial carcinoma with cervical stromal involvement overt in clinical examination or confirmed by endocervical curettage). 2009. (Accessed at http://www.acrin.org/TabID/156/Default.aspx.)

5. ACRIN Protocol #6685: A multicenter trial of FDG-PET/CT staging of head and neck cancer and its impact on the N0 neck surgical treatment in head and neck cancer patients., 2005. (Accessed at http://www.acrin.org/TabID/569/Default.aspx.)

6. ACRIN Protocol #6678: FDG-PET/CT as a predictive marker of tumor response and patient outcome: Prospective validation in non-small cell lung cancer. 2010. (Accessed at http://www.acrin.org/TabID/162/Default.aspx.)

7. ACRIN Protocol #6665: A phase II trial of neoadjuvent /adjuvant STI-571 (Gleevec NSC #716051) for primary and recurrent operable malignant GIST expressing the KIT receptor tyrosine kinase (CD117). 2002. (Accessed at http://www.acrin.org/TabID/152/Default.aspx.)

8. ACRIN Imaging Core Laboratory Standard Operating Procedure: PET Scanner Qualification Image Review and Quality Control Analysis. 2010. (Accessed at http://www.acrin.org/Portals/0/Corelabs/ACRINPETScannerQual_SOP 3-17-10.pdf.)

9. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM Procedure Guidelines for Tumour PET/CT Imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2014 in press.

10. FDG-PET/CT Technical Committee. FDG-PET/CT as an Imaging Biomarker Measuring Response to Cancer Therapy Profile, Quantitative Imaging Biomarkers Alliance. Version 1.05. Publicly Reviewed Version. QIBA, December 11, 2013. Available from: RSNA.ORG/QIBA.

11. Boellaard R, Oyen WJ, Hoekstra CJ, et al. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. Eur J Nucl Med Mol Imaging 2008;35:2320-33.

12. Hallett WA, Maguire RP, McCarthy TJ, Schmidt ME, Young H. Considerations for generic oncology FDG-PET/CT protocol preparation in drug development. IDrugs 2007;10:791-6.

13. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. J Clin Oncol 2007;25:579-86.

14. SNM Clinical Trials Network Summit: World-wide Standardization of FDG PET Imaging Protocols for Clinical Trials. In. Salt Lake City, Utah; 2011.

15. de Langen AJ, Vincent A, Velasquez LM, et al. Repeatability of 18F-FDG uptake measurements in tumors: a metaanalysis. J Nucl Med 2012;53:701-8.

16. Delbeke D, Coleman RE, Guiberteau MJ, et al. Procedure guideline for tumor imaging with 18F-FDG PET/CT 1.0. J Nucl Med 2006;47:885-95.

17. Coleman RE, Delbeke D, Guiberteau MJ, et al. Concurrent PET/CT with an integrated imaging system: intersociety dialogue from the joint working group of the American College of Radiology, the Society of Nuclear Medicine, and the Society of Computed Body Tomography and Magnetic Resonance. J Nucl Med 2005;46:1225-39.

18. ACRIN Imaging Core Laboratory Standard Operating Procedure: Patient Preparation and FDG Administration. 2009. (Accessed at http://www.acrin.org/Portals/0/Corelabs/ACRINPatientPrep_SOP 3-17-10.pdf.)

19. Lindholm H, Johansson O, Jonsson C, Jacobsson H. The distribution of FDG at PET examinations constitutes a relative mechanism: significant effects at activity quantification in patients with a high muscular uptake. Eur J Nucl Med Mol Imaging 2012;39:1685-90.

20. Williams G, Kolodny GM. Suppression of myocardial 18F-FDG uptake by preparing patients with a high-fat, low-carbohydrate diet. AJR Am J Roentgenol 2008;190:W151-6.

21. Wykrzykowska J, Lehman S, Williams G, et al. Imaging of inflamed and vulnerable plaque in coronary arteries with 18F-FDG PET/CT in patients with suppression of myocardial uptake using a low-carbohydrate, high-fat preparation. J Nucl Med 2009;50:563-8.

22. Soderlund V, Larsson SA, Jacobsson H. Reduction of FDG uptake in brown adipose tissue in clinical patients by a single dose of propranolol. Eur J Nucl Med Mol Imaging 2007;34:1018-22.

23. Agrawal A, Nair N, Baghel NS. A novel approach for reduction of brown fat uptake on FDG PET. Br J Radiol 2009;82:626-31.

24. Treves ST, Parisi MT, Gelfand MJ. Pediatric radiopharmaceutical doses: new guidelines. Radiology 2011;261:347-9.

25. Gelfand MJ, Parisi MT, Treves ST. Pediatric radiopharmaceutical administered doses: 2010 North American consensus guidelines. J Nucl Med 2011;52:318-22.

26. Hamberg LM, Hunter GJ, Alpert NM, et al. The dose uptake ratio as an index of glucose metabolism: useful parameter or oversimplification? J Nucl Med 1994;35:1308-12.

27. Lodge MA, Lucas JD, Marsden PK, et al. A PET study of 18FDG uptake in soft tissue masses. Eur J Nucl Med 1999;26:22-30.

28. ACRIN Imaging Core Laboratory Standard Operating Procedure: Image Acquisition and Processing for ACRIN PET Scans. 2009. (Accessed at http://www.acrin.org/Portals/0/Corelabs/ACRINAcquisition_SOP923_5-05-07_.pdf.)

29. Beyer T, Antoch G, Muller S, et al. Acquisition protocol considerations for combined PET/CT imaging. *J Nucl Med* 2004;45 Suppl 1:25S-35S.

30. Boellaard R. Need for standardization of 18F-FDG PET/CT for treatment response assessments. J Nucl Med 2011;52 Suppl 2:93S-100S.

31. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med 2009;50 Suppl 1:122S-50S.

32. James W. Research on Obesity. London; Her Majesty's Stationery Office, 1976

33. Tahari AK, Chien D, Azadi JR, Wahl RL. Optimum lean body formulation for correction of standardized uptake value in PET imaging. J Nucl Med. 2014 Sep;55(9):1481-4.

34. Janmahasatian S, Duffull SB, Ash S, et al. Quantification of lean bodyweight. Clin Pharmacokinet 2005;44:1051-65.

35. Morgan DJ, Bray KM. Lean body mass as a predictor of drug dosage. Implications for drug therapy. *Clin Pharmacokinet* 1994;26:292-307.

36. Sugawara Y, Zasadny KR, Neuhoff AW, Wahl RL. Reevaluation of the standardized uptake value for FDG: variations with body weight and methods for correction. *Radiology* 1999;213:521-5.

37. Hanley MJ, Abernethy DR, Greenblatt DJ. Effect of obesity on the pharmacokinetics of drugs in humans. *Clin Pharmacokinet* 2010;49:71-87.

38. Chan T. Computerized method for automatic evaluation of lean body mass from PET/CT: comparison with predictive equations. J Nucl Med 2012;53:130-7.

39. Lodge MA, Chaudhry MA, Wahl RL. Noise considerations for PET quantification using maximum and peak standardized uptake value. J Nucl Med 2012;53:1041-7.

40. Velasquez LM, Boellaard R, Kollia G, et al. Repeatability of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies. J Nucl Med 2009;50:1646-54.

41. ACRIN Imaging Core Laboratory Standard Operating Procedure: Procedure for physician analysis of PET and PET/CT images for ACRIN quantitative PET protocols. 2007. (Accessed at http://www.acrin.org/Portals/0/Corelabs/ACRINPhysicianAnalysis_SOP924_5-05-07_.pdf.)

42. Larson SM, Erdi Y, Akhurst T, et al. Tumor Treatment Response Based on Visual and Quantitative Changes in Global Tumor Glycolysis Using PET-FDG Imaging. The Visual Response Score and the Change in Total Lesion Glycolysis. Clin Positron Imaging 1999;2:159-71.

43. Paquet N, Albert A, Foidart J, Hustinx R. Within-patient variability of (18)F-FDG: standardized uptake values in normal tissues. J Nucl Med 2004;45:784-8.

44. Huang SC. Anatomy of SUV. Standardized uptake value. Nucl Med Biol 2000;27:643-6.

45. E.A. Eisenhauera, P. Therasse, J. Bogaerts, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1) Eur J Cancer, 45 (2) (2009 Jan), pp. 228–247

46. Clarke JL, Chang S. Pseudoprogression and pseudoresponse: challenges in brain tumor imaging. Curr Neurol Neurosci Rep 2009;9:241-6.

47. Hygino da Cruz LC, Jr., Rodriguez I, Domingues RC, Gasparetto EL, Sorensen AG. Pseudoprogression and pseudoresponse: imaging challenges in the assessment of posttreatment glioma. AJNR Am J Neuroradiol 2011;32:1978-85.

48. Caroline I, Rosenthal MA. Imaging modalities in high-grade gliomas: pseudoprogression, recurrence, or necrosis? J Clin Neurosci 2012;19:633-7.

49. Oborski MJ, Laymon CM, Lieberman FS, Mountz JM. Distinguishing Pseudoprogression From Progression in High-Grade Gliomas: A Brief Review of Current Clinical Practice and Demonstration of the Potential Value of 18F-FDG PET. Clin Nucl Med 2013.

50. Scheuermann JS, Saffer JR, Karp JS, Levering AM, Siegel BA. Qualification of PET scanners for use in multicenter cancer clinical trials: the American College of Radiology Imaging Network experience. J Nucl Med 2009;50:1187-93.

51. SNMMI Clinical Trials Network web site: http://interactive.snm.org/index.cfm?PageID=10641

52. Alessio AM, Kinahan PE, Manchanda V, et al. Weight-based, low-dose pediatric whole-body PET/CT protocols. J Nucl Med 2009;50:1570-7.

53. Huang B, Law MW, Khong PL. Whole-body PET/CT scanning: estimation of radiation dose and cancer risk. Radiology 2009;251:166-74.