FDG-PET Subcommittee Meeting

May 19, 2009

Intersection between medical need and what is technically possible. Relationship to effect size.

To compare effect size, need two different populations, treated and untreated, and must be considered in the context of intended use.

10% reproducibility is a reasonable goal to achieve for multi-site studies.

What is better—lower cost; smaller 'n'; clinically relevant

Assume you are imaging the same patient in the same scanner for now; ideally you would get the same value from different machines in different centers.

Need to get SUV's on a normal person to use ratios to correct for calibration.

Longitudinal quantitation* of tumor metabolism via FDG-PET/CT that can be used practically and efficiently as a biomarker, with known precision and accuracy, in multicenter (i.e., multi-vendor) clinical trials and meta-analyses, enhancing its potential for use in patient care and qualification for use in clinical studies of pathophysiology and therapeutic interventions.

*measuring change over time (e.g., pre- and post-treatment in a clinical trial of an investigational agent or for refinement of patient care during a course of therapy.

[Footnote: This may include meta-analysis across trials and databases and may be more broadly applicable to other "final common pathway" tracers of the hallmarks of cancer (e.g., proliferation, apoptosis, angiogenesis, etc), any of which would improve confidence around expanding the indications of PET/CT in response monitoring for both clinical trials and patient care.]

Need for prioritized request to vendors.

Quantification includes proper calibration and resolution (partial volume effect) but will be tackled later, along with other issues from the original brainstorming session. Specification of performance targets might be based on pharma needs to reduce the number of patients required in clinical trials for a given effect size. In order to do this, it is necessary for vendors to measure and report performance metrics for SUV.

Develop claims: Rationale, limitations, proposed solutions, prioritized.

Claim #4: Can find in the DICOM header reliable data required for SUV calculation

Precursors:

- vendors have supplied DICOM header info and conformance statements(untrusted)
- samples (test/retest) actual images of NEMA phantom from multiple vendors and systems
- understanding of physics to produce synthetic phantom
- experience with use of DROs for reconstruction scheme evaluation

Rationale:

- is no existing DRO for QC of sites/systems
- is no way to readily compare SUV from different analysis software
- *physical phantoms are not sufficient to solve this class of problem, even though they obviate the need to continually update a DRO for each new product version*

Solution:

DRO group to merge with SUV calculation group, because cannot validate calculation based on header without pixel data source and known expected values

Short term (threshold) goal – single standard stack (or one for each vendor based on what we know).

Middle term (target) goal – validate receiving workstations with vendor-specific stack with header (and standardized SUV truth values) – supply floating point data, abstract meta-data, vendor produces pixel data and DICOM header and series organization – if vendor can provide multiple forms (e.g., in SUVbw or MBq/ml or PROPCNTS) provide all of them (overlap with s/w version group – needs to tell us what changes and how it affects computation) – supply pre-defined ROIs with pre-defined truth values for them – need to define supplied slice thickness, FOV and resolution, including cross-plane for 3D SUV calculation (i.e, on coronal or sagital recons or using a 3D (cylindrical or spherical)) – CT data NOT needed *unless* want to validate regional measurements as opposed to point measurements (a refinement to the middle term (target) goal) - want two phantoms – one with no noise so that the values are absolutely predictable, and with noise for

realism (want assymetric data for regional calculation check). NOT A PRODUCT FEATURE BUT SAMPLE DATA ONLY.

Long term (ideal) goal – validate image generation beyond acquisition point AND receiving workstation - reconstruct from standard raw data (with provided reconstructed attenuation (not diagnostic) CT image), prior to any corrections (sinugram, list mode) - measures software performance ONLY, not hardware problems (non-goal – need real phantoms to detect such changes); CT is required – need to define corresponding slice thickness, FOV and resolution (non-goal is testing registration fusion, which requires a different type)

Use NEMA NU2 phantom as basis for synthetic phantom data.

Additional use-cases:

- feeding other vendors consoles with their own phantom data (e.g., from different versions of their own hardware) ? What about other vendors' data on a vendor's console ?
- what about phantoms to validate registration, and the impact on SUV, especially deformable registration

Include software version info ?

Include calibration info ?

1st, measure (accurate and precise)

2nd, change metric (longitudinal)

3rd, response metric (clinical relevance)

Quantification includes proper calibration and resolution (partial volume effect) but will be tackled later, along with other issues from the original brainstorming session. Specification of performance targets might be based on pharma needs to reduce the number of patients required in clinical trials for a given effect size. In order to do this, it is necessary for vendors to measure and report performance metrics for SUV.