QIBA Profile. Computed Tomography: Change Measurements in the Volumes of Solid Tumors

Version 2.0

28 July 2011

Table of Contents

Open Issues:

The following issues have not been resolved to the satisfaction of the technical committee. An open issue may be a short question prompting a proposed resolution or discussion. The issues and answers below may represent some of the directions the Committee is currently leaning. Feedback on these issues is encouraged, particularly during the Public Comment period for the profile.

Q. Is the claim appropriate/supported by the profile details, published literature, and QIBA groundwork? Is it stated in clear and statistically appropriate terms? A.
Q. What kind of additional study (if any is needed) would best prove the profile claim? A.
Q. How do we balance specifying what to accomplish vs how to accomplish it? A. E.g. if the requirement is that the scan be performed the same way, do we need to specify that the system or the Technologist record how each scan is performed? If we don't, how will the requirement to "do it the same" be met?
Q. Should there be a "patient appropriateness" or "subject selection" section? A. The protocol template includes such a section to describe characteristics of appropriate (and/or inappropriate) subjects. E.g. a requirement that the patient be able to hold their breath for 15 seconds. We could also discuss what constitutes an "assessable lesion" (the claim introduces this term)
Q. Does 4cm/sec "scan speed" preclude too many sites? A. A 4cm /sec threshold would likely forestall a lot of potential breath hold issues.
Q. What do we mean by noise and how do we measure it? A.
Q. Is 5HU StdDev a reasonable noise value for all organs? A. If it's not, should we allow multivalued specifications for different organs/body regions? Should we simply have several profiles?
Q. Are there sufficient DICOM fields for all of what we need to record in the image header, and what are they specifically? A. For those that exist, we need to name them explicitly. For those that may not currently exist, we need to work with the appropriate committees to have them added.
Q. Have we worked out the details for how we establish compliance to these specifications? A. We are continuing to work on how this is to be accomplished but felt that it was helpful to start the review process for the specifications in parallel with working on the compliance process.
Q. What is the basis of the specification of 15% for the variability in lesion volume assessment

within the Image Analysis section, and is it inclusive or exclusive of reader performance?
A. As stated it is inclusive of reader performance, with a view to be consistent with the overall claim and where this action takes place in the pipeline process. We acknowledge that allocation of variability across the chain is fraught with difficulty and also that accounting for reader performance is also difficult in the presence of different levels of training and competence among readers. Input on these points to help with this is appreciated (as is also the case for all aspects of this Profile).

Closed Issues:

The following issues have been considered closed by the technical committee. They are provided here to forestall discussion of issues that have already been raised and resolved, and to provide a record of the rationale behind the resolution.

Q. Should we specify all three levels (Acceptable, Target, Ideal) for each parameter?
A. No. As much as possible, provide just the Acceptable value. The Acceptable values should be selected such that the profile claim will be satisfied.

Q. What is the basis for our claim, and is it only aspirational?
A. Our claim is informed by an extensive literature review of results achieved under a variety of conditions. From this perspective it may be said to be well founded; however, we acknowledge that the various studies have all used differing approaches and conditions that may be closer or farther from the specification outlined in this document. In fact the purpose of this document is to fill this community need. Until field tested, the claim may be said to be “consensus.” Commentary to this effect has been added in the Claims section, and the Background Information appendix has been augmented with the table summarizing our literature sources.

Q. What about dose?
A. A discussion has been added to address dose issues. Increased radiation absorbed dose improves SNR and gives better lesion definition up to a point.

I. Executive Summary

X-ray computed tomography provides an effective imaging technique for assessing treatment response in patients with cancer. Quantification is helpful when tumor masses change relatively slowly over the course of illness. Currently most size measurements are uni-dimensional estimates of longest diameters (LDs) on axial slices, as specified by RECIST (Response Evaluation Criteria In Solid Tumors). Since its introduction, limitations of this method have been reported. Many investigators have suggested that quantifying whole tumor volumes could solve some of the limitations of depending on diameter measures, and may have a major impact on patient management [1-2]. An increasing number of studies have shown that volumetry has value [3-12].

QIBA has constructed a systematic approach for standardizing and qualifying volumetry as a biomarker of response to treatments for a variety of medical conditions, including cancers in the lung (either primary cancers or cancers that metastasize to the lung [18]). Several studies with varying scope are now underway to provide comparison between the effectiveness of volumetry and uni-dimensional LDs as the basis for RECIST in multi-site, multi-scanner-vendor settings. This QIBA Profile is expected to provide specifications that may be adopted by users as well as equipment developers to meet targeted levels of clinical performance in identified settings.

This profile makes claims about the precision with which changes in tumor volumes can be measured under a set of defined image acquisition, processing, and analysis conditions.

The intended audiences include:

- Technical staffs of software developers and device manufacturers who create products for this purpose
- Clinical trial scientists and physician PIs of clinical trials
- Practicing clinicians at healthcare institutions considering appropriate specifications for procuring new equipment
- Experts involved in quantitative medical image analysis
- Anyone interested in the technical and clinical aspects of medical imaging

Note that specifications stated as “requirements” here are only requirements to achieve the claim, not “requirements on standard of care.” Specifically, meeting the goals of the profile are secondary to properly caring for the patient.

II. Clinical Context and Claims

Utilities and Endpoints for Clinical Trials

These specifications are appropriate for quantifying the volumes of malignant lesions and measuring their longitudinal changes within subjects. The primary objective is to evaluate their growth or regression with serially acquired CT scans and image processing techniques.

Compliance with this profile by relevant staff and equipment supports the following claim(s):

Claim: Measure Change in Tumor Volume

Increases or decreases of more than 30% in a tumor's volume measured over time is above the measurement variability and associated with a true biological change given that the tumor is measurable (i.e., tumor margins should be recognizable on all images in both scans), and the longest diameter of the tumor is 10 mm or greater in the initial scan. This means that technical variation in the measurement is no more than 15% (half of the 30% claimed for biological significance).

This claim has been informed by an extensive review of the literature, as summarized in the Background Information appendix. It is currently a consensus claim that has not yet been fully substantiated by studies that strictly conform to the specifications given here. To date there has not existed a standard utilized by a sufficient number of studies. The expectation is that during field test, data on the actual field performance will be collected and changes made to the claim or the details accordingly. At that point, this caveat may be removed or re-stated.

III. Profile Details

A technical description of tests for the biomarker, identifying measurement activities and read-outs, is provided:

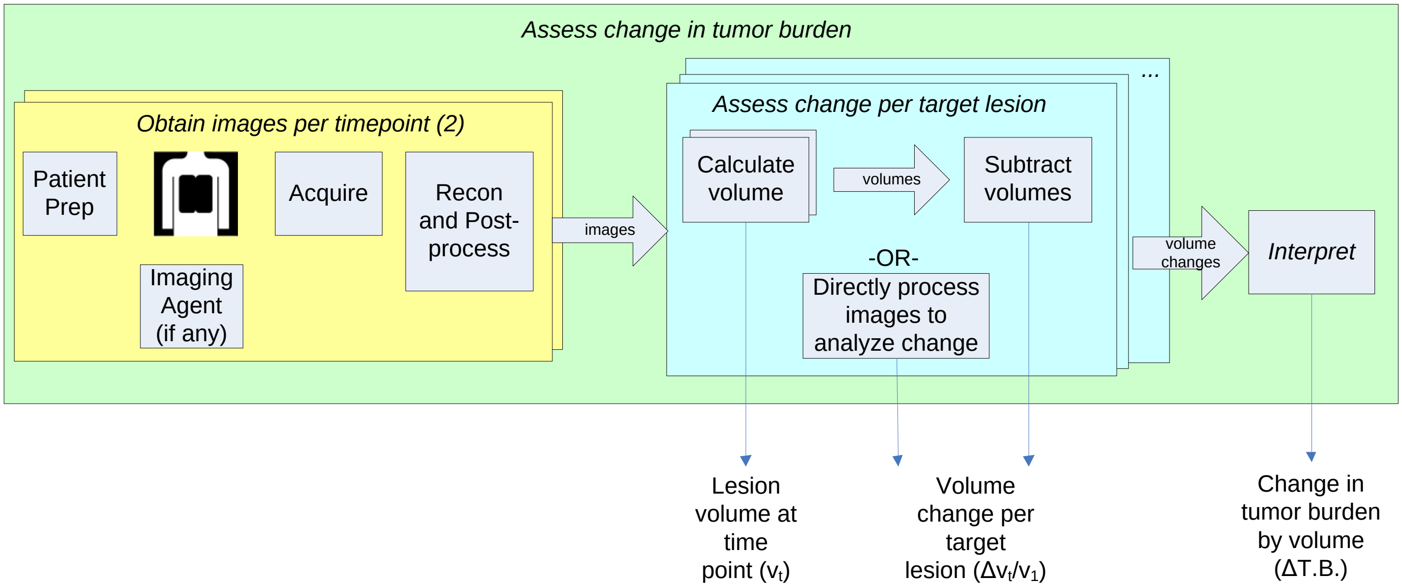

Figure 1: The assay method for computing and interpreting volumetric assessment using computed tomography may be described as a pipeline. Patients (or subjects) are prepared for scanning, a imaging agent to enhance contrast may or may not be used, raw image data is acquired, and images are formed using mathematical reconstruction and/or post-processing methods. Images may be obtained at a multiplicity of time points, notably at two time points for a change assessment as is considered by this document. Images formed at each of the two time points serve as the input to the downstream image analysis activity to assess the degree of change per each target lesion. Detection of target lesions as well as classification as to whether they are target and/or evaluable lesions is beyond the scope of this document. For each detected and evaluable target lesion, change may be assessed by calculating absolute volume at each of the two time points and performing a subtraction, or alternatively through other means that may be proposed wherein a direct measure of change is assessed

without specific regard to the absolute volumes. Philosophically it is desired that the profile encourage rather than discourage innovation in the means by which this is done, however, in the end the change is assessed as a percentage according to the formula (delta in volume between the two time points)/volume at time point 1. Downstream from this analysis the change may be interpreted according to a variety of different response criteria. These response criteria are beyond the scope of this document.

Formally defined "Actors" who are required to meet these claims include the following:

- Hardware and software devices (acquisition, reconstruction, and analysis)
- Technologists
- Image Analysts
- Image Acquisition Sites

The following sections provide details for what the various components required for compliance:

Section 1, Subject Handling, is practiced by an Image Acquisition Site.

Section 2, Imaging Data Acquisition, is practiced by a Technologist at an Image Acquisition Site using an Acquisition Device.

Section 3, Imaging Data Reconstruction, is practiced by an Technologist at an Image Acquisition Site using Reconstruction Software.

Section 4, Image Analysis, is practiced by an Image Analyst using one or more Software Analysis Tools.

The requirements included herein are intended to establish a baseline level of capabilities. Providing higher performance or advanced capabilities is both allowed and encouraged. The profile is not intended to be limiting in any way with respect to how these requirements are met by equipment suppliers.

Note that this profile is "lesion-oriented", meaning that different lesions in different anatomic regions might be imaged and processed with different parameters as long as any given lesion is handled the same way each time.

1. Subject Handling

1.1 Timing Relative to Index Intervention Activity

The pre-treatment CT scan shall take place prior to any intervention to treat the disease. This scan is referred to as the "*baseline*" scan. It should be acquired as soon as possible before the initiation of treatment, and in no case more than the number of days before treatment specified in the protocol.

1.2 Timing Relative to Confounding Activities

This document does not presume any timing relative to other activities. Fasting prior to a contemporaneous FDG PET scan or the administration of oral contrast for abdominal CT are not expected to have any adverse impact on this profile.

1.3 Contrast Preparation and Administration

DISCUSSION

The use of contrast is not an absolute requirement for this profile. However, the use of contrast material (intravenous or oral) may be medically indicated in defined clinical settings. Contrast characteristics influence the appearance, conspicuity, and quantification of tumor volumes.

SPECIFICATION

Parameter	**Specification**
Use of intravenous or oral contrast	The Technologist shall use equivalent contrast (including dose calculation, schedule, administration route, and rate) as used at baseline for subsequent time points. If not used at baseline, it shall not be used in follow-up scans.
Image Header	The Acquisition Device shall record the use and type of contrast, actual dose calculation, schedule rate, delay, and apparatus utilized in the image header. This may be by automatic interface with contrast administration devices in combination with text entry fields that shall be filled in by the Technologist.

1.4 Subject Positioning

DISCUSSION

Consistent positioning avoids unnecessary variance in attenuation, changes in gravity induced shape and fluid distribution, or changes in anatomical shape due to posture, contortion, etc. Significant details of subject positioning include the position of their upper extremities, the anterior-to-posterior curvature of their spines as determined by pillows under their backs or knees, the lateral straightness of their spines, and, if prone, the direction the head is turned. Positioning the subject Supine/Arms Up/Feet first has the advantage of promoting consistency, and reducing cases where intravenous lines go through the gantry, which could introduce artifacts.

SPECIFICATION

Parameter	**Specification**
Subject Positioning	The Technologist shall position the subject the same as for prior scans. If the previous positioning is unknown, the Technologist shall position the subject Supine/Arms Up/Feet first if possible.
Table Height	The Technologist shall adjust the table height to place the mid-axillary line at isocenter.
Image Header	The Acquisition Device shall record the Table Height in the image header.

1.5 Instructions to Subject During Acquisition

DISCUSSION

Breath holding reduces motion that might degrade the image. Full inspiration inflates the lungs, which separates structures and makes lesions more conspicuous.

Although performing the acquisition in several segments (each of which has an appropriate breath hold state) is possible, performing the acquisition in a single breath hold is likely to be more easily repeatable and does not depend on the Technologist knowing where the lesions are located.

SPECIFICATION

Parameter	**Specification**
Breath hold	The Technologist shall ensure that image acquisition occurs at least near the high end inspiration. The Technologist shall ensure that for each lesion the breath hold state is the same as for prior scans.
Image Header	The Technologist shall record factors that adversely influence patient positioning or limit their ability to cooperate (e.g., breath hold, remaining motionless, agitation in patients with decreased levels of consciousness, patients with chronic pain syndromes, etc.). These shall be accommodated with data entry fields provided by the Acquisition Device.

1.6 Timing/Triggers

DISCUSSION

The amount and distribution of contrast at the time of acquisition can affect the appearance and conspicuity of lesions.

SPECIFICATION

Parameter	**Specification**
Timing / Triggers	The Technologist shall ensure that the time-interval between the administration of intravenous contrast (or the detection of bolus arrival) and the start of the image acquisition is the same as for prior scans.
Image Header	The Acquisition Device shall record actual Timing and Triggers in the image header.

2. Image Data Acquisition

DISCUSSION

CT scans for tumor volumetric analysis will be performed on equipment that complies with the specifications set out in this profile. At this stage of development, we continue to recommend that all CT scans for an individual participant be performed on the same platform throughout the trial. In the rare instance of equipment malfunction, follow-up scans on an individual participant can be performed on the same type of platform. All efforts should be made to have the follow-up scans performed with identical parameters as the first. This is inclusive of as many of the scanning parameters as possible, including the same field of view (FOV).

A set of scout images should be initially obtained. Pitch is chosen so as to allow completion of the scan in a single breath hold. In some cases two or more breaths may be necessary. In those cases, it is important that the target lesion be fully included within one of the sequences.

Faster scans shorten the scan time and reduce the breath hold requirements, thus reducing the likelihood of motion artifacts. Scan Plane (transaxial is preferred) may differ for some subjects due to the need to position for physical deformities or external hardware.

Total Collimation Width (defined as the total nominal beam width) is often not directly visible in the scanner interface. Wider collimation widths can increase coverage and shorten acquisition, but can introduce cone beam artifacts which may degrade image quality.

Slice Width directly affects voxel size along the subject z-axis. Smaller voxels are preferable to reduce partial volume effects and provide higher accuracy due to higher spatial resolution.

X-ray CT uses ionizing radiation. Exposure to radiation can pose risks. It is recognized that there are tradeoffs between radiation dose and image quality. As the radiation dose is reduced, image quality can be degraded. It is expected that health care professionals will balance the need for good image quality with the risks of radiation exposure on a case-by-case basis. It is not within the scope of this document to describe how these trade-offs should be resolved.

SPECIFICATION

Parameter	**Specification**
Scan Duration for Thorax	The Acquisition Device shall be capable of performing the required scans at an axial rate of at least 4cm per second.
Anatomic Coverage	The Technologist shall perform the scan such that the acquired anatomy is the same as for prior scans.
Scan Plane (Image Orientation)	The Technologist shall set the scan plane to be the same as for prior scans.
Total Collimation Width	The Acquisition Device shall be set up so as to achieve a total collimation width >=20mm.
IEC Pitch	The Acquisition Device shall be set up so as to achieve IEC pitch less than 1.5.
Tube Potential	The Acquisition Device shall be set up so as to achieve same kVp for all scans
Single Collimation Width	The Acquisition Device shall be set up so as to achieve single collimation width <= 1.5mm.
Image Header	The Acquisition Device shall record actual Anatomic Coverage, Field of View, Scan Duration, Scan Plane, Total Collimation Width, Single Collimation Width, Scan Pitch,

Parameter	Specification
	Tube Potential, and Slice Width in the image header.

3. Image Data Reconstruction

DISCUSSION

It is acknowledged that image reconstruction is closely related to image acquisition. These specifications are the result of discussions to allow a degree of separation in their consideration without suggesting they are totally independent.

Spatial Resolution quantifies the ability to resolve spatial details. Lower spatial resolution can make it difficult to accurately determine the borders of tumors, and as a consequence, decreases the precision of volume measurements. Increased spatial resolution typically comes with an increase in noise. Therefore, the choice of factors that affect spatial resolution typically represent a balance between the need to accurately represent fine spatial details of objects (such as the boundaries of tumors) and the noise within the image. Spatial resolution is mostly determined by the scanner geometry (which is not usually under user control) and the reconstruction kernel (which is somewhat under user control as the user usually gets to choose from a limited set of choices of reconstruction kernels provided at the scanner). It is stated in terms of “the number of line-pairs per cm that can be resolved in a scan of resolution phantom (such as the synthetic model provided by the American College of Radiology and other professional organizations).” –OR– “the full width at half of the line spread function”.

Noise Metrics quantify the magnitude of the random variation in reconstructed CT numbers. Some properties of the noise can be characterized by the standard deviation of reconstructed CT numbers over a uniform region in phantom. Noise (pixel standard deviation) can be reduced by using thicker slices for a given mAs. A constant value for the noise metric might be achieved by increasing mAs for thinner slices and reducing mAs for thicker slices. The standard deviation is limited since it can vary by changing the reconstruction kernel, which will also impact the spatial resolution. A more comprehensive metric would be the noise-power spectrum which measures the noise correlation at different spatial frequencies.

Reconstruction Field of View affects reconstructed pixel size because the fixed image matrix size of most CT scanners is 512 X 512. If it is necessary to expand the field of view to encompass more anatomy, the resulting larger pixels may be insufficient to achieve the claim. A targeted reconstruction with a smaller field of view may be necessary, but a reconstruction with that field of view would need to be performed for every time point. Pixel Size directly affects voxel size along the subject x-axis and y-axis. Smaller voxels are preferable to reduce partial volume effects and provide higher measurement precision. Pixel size in each dimension is not the same as resolution in each dimension; inherent resolution is different than how the data is reconstructed and is strongly affected by the reconstruction kernel.   When comparing data fields of different resolution, do not sacrifice higher resolution data to match the level of lower resolution data.

Reconstruction Interval (a.k.a. Slice spacing) that results in discontiguous data is unacceptable as they may “truncate” the spatial extent of the tumor, degrade the identification of tumor boundaries, confound the precision of measurement for total tumor volumes, etc. Decisions about overlap (having an interval that is less than the nominal reconstructed slice thickness) need to consider the technical requirements of the clinical trial, including effects on measurement, throughput, image analysis time, and storage requirements.

Reconstructing datasets with overlap will increase the number of images and may slow down throughput, increase reading time and increase storage requirements. For multidetector row CT (MDCT) scanners, creating overlapping image data sets has NO effect on radiation exposure; this is true because multiple reconstructions having different kernel, slice thickness and intervals can be reconstructed from the same acquisition (raw projection data) and therefore no additional radiation exposure is needed. <Note that the slice thickness is "nominal" since the thickness is not technically the same at the middle and the edges>

Reconstruction Kernel Characteristics need to be defined to optimize the analysis for each lesion while still meeting the requirements for noise and spatial resolution. A softer kernel can reduce noise at the expense of spatial resolution. An enhancing kernel can improve resolving power at the expense of increased noise.

The effects of iterative reconstructions on quantitative accuracy and reproducibility are currently not fully understood as of this writing of this profile version.

SPECIFICATION

For quantification of whole tumor volumes, the reconstruction software produces images that meet the following specifications:

Parameter	**Specification**
Spatial Resolution	The Reconstruction Software shall be set up so as to achieve spatial resolution >= 6 lp/cm –OR– Axial FWHM <= 0.8mm.
Voxel Noise	The Reconstruction Software shall be set up so as to achieve voxel noise standard deviation of < 5HU in 20cm water phantom.
Reconstruction Field of View	The Reconstruction Software shall be set up so as to achieve a reconstruction field of view spanning the entire lateral extent of the patient, but no greater than required to image the entire body; <same as previous scan>
Slice Thickness	The Reconstruction Software shall be set up so as to achieve slice thickness ≤2.5 mm.
Reconstruction Interval	The Reconstruction Software shall be set up so as to achieve reconstruction interval ≤2.5 mm.
Reconstruction Overlap	The Reconstruction Software shall be set up so as to achieve reconstruction overlap > 0 (i.e. no gap, and may have some overlap).
Reconstruction Kernel Characteristics	The Reconstruction Software shall be set up so as to utilize an equivalent kernel for all time points.
Image Header	The Reconstruction Software shall record actual Spatial Resolution, Noise, Pixel Spacing, Reconstruction Interval, Reconstruction Overlap, Reconstruction Kernel Characteristics, as well as the model-specific Reconstruction Software parameters utilized to achieve compliance with these metrics in the image header.

4. Image Analysis

DISCUSSION

Each lesion is characterized by determining the boundary of the lesion (referred to as segmentation), then computing the volume of the segmented lesion. Segmentation may be performed automatically by a software algorithm, manually by a human observer, or semi-automatically by an algorithm working with human guidance/intervention. The volume of the segmented region is then computed automatically from the segmented boundary. Many Analysis Software Tools segment various types of tumors on CT images based on a starting seed point, stroke, or region and change is assessed as the difference of two volume computations. It is acknowledged that computing absolute volumes at two separate time points is only one way to approach the change calculation. Methods that calculate volume changes directly without calculating volumes at individual time points are acceptable so long as the results are compliant with these specifications as set out by this profile.

SPECIFICATION

Parameter	**Specification**
Common Lesion Selection	The Image Analysis Tool shall allow a common set of lesions to be designated for measurement, which are then subsequently measured by all readers.
Lesion Volume Change	The Image Analysis Tool shall measure lesion volume change (according to Figure 1) with variability less than +/- 15%.
Multiple Lesions	The Image Analysis Tool shall allow multiple lesions to be measured, and each measured lesion to be associated with a human-readable identifier that can be used for correlation across time points.
Recording	The Image Analysis Tool shall record actual model-specific Analysis Software set-up and configuration parameters utilized to achieve compliance with these metrics shall be recorded. Image Analysis Tools shall record in (and reload for review from) region specification (e.g., lesion segmentation boundary) and volumetric measurement as well as metadata in standard formats including one or more of the following output formats: DICOM Presentation State, DICOM Structured Report; DICOM RT Structure Set; DICOM raster or surface segmentation.

IV. Compliance

Acquisition Device

Compliance is certified according to specifications set out in the Image Acquisition section above. Additionally, compliant Acquisition Devices shall provide means to record the information identified in the Subject Handling section as means to document compliance of the Image Acquisition Site to the specifications noted there.

Reconstruction Software

Compliance to specifications as set out in the Image Reconstruction section above. Additionally, compliant Reconstruction Software shall propagate the information collected at the prior Subject Handling and

Imaging Acquisition stages and extend it with those items noted in the Reconstruction section. See the compliance procedure notes associated with Acquisition Devices above for procedural assistance to identify Model Specific Parameters for Reconstruction Software.

Software Analysis Tool

Compliance to specifications as set out in the Image Analysis section above. Additionally, compliant Software Analysis Tools shall propagate the information collected at the prior Subject Handling, Imaging Acquisition, and Imaging Reconstruction stages and extend it with those items noted in the Analysis section

Image Acquisition Site

Typically clinical sites are selected due to their competence in oncology and access to a sufficiently large patient population under consideration. For imaging it is important to consider the availability of:

- appropriate imaging equipment and quality control processes,
- appropriate injector equipment and contrast media,
- experienced CT Technologists for the imaging procedure, and
- processes that assure imaging profile compliant image generation at the correct point in time.

A calibration and QA program shall be designed consistent with the goals of the clinical trial. This program shall include (a) elements to verify that sites are performing correctly, and (b) elements to verify that sites' CT scanner(s) is (are) performing within specified calibration values. These may involve additional phantom testing that address issues relating to both radiation dose and image quality (which may include issues relating to water calibration, uniformity, noise, spatial resolution -in the axial plane-, reconstructed slice thickness z-axis resolution, contrast scale, CT number calibration and others). This phantom testing may be done in additional to the QA program defined by the device manufacturer as it evaluates performance that is specific to the goals of the clinical trial.

References

[1] Moertel CG, Hanley JA. The effect of measuring error on the results of therapeutic trials in advanced disease. Disease 1976; 38: 388-394.

[2] Quivey JM, Castro JR, Chen GT, Moss A, Marks WM. Computerized tomography in the quantitative assessment of tumour response. Br J Disease Suppl 1980; 4:30-34.

[3] Munzenrider JE, Pilepich M, Rene-Ferrero JB, Tchakarova I, Carter BL. Use of body scanner in radiotherapy treatment planning. Disease 1977; 40:170-179.

[4] Wormanns, D., Kohl, G., Klotz, E., Marheine, A., Beyer, F., Heindel, W., and Diederich, S. Volumetric measurements of pulmonary nodules at multi-row detector CT: In vivo reproducibility. Eur Radiol 14: 86–92, 2004.

[5] Kostis WJ, Yankelevitz DF, Reeves AP, Fluture SC, Henschke CI, Small Pulmonary Nodules: Reproducibility of Three-dimensional Volumetric Measurement and Estimation of Time to Follow-up CT, Radiology, Volume 231 Number 2, 2004.

[6] Revel M-P, Lefort C, Bissery A, Bienvenu M, Aycard L, Chatellier G, Frija G, Pulmonary Nodules: Preliminary Experience with Three-dimensional Evaluation, Radiology May 2004.

[7] Marten K, Auer F, Schmidt S, Kohl G, Rummeny EJ, Engelke C, Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria, Eur Radiol (2006) 16: 781–790.

[8] Goodman, L.R., Gulsun, M., Washington, L., Nagy, P.G., and Piacsek, K.L. Inherent variability of CT lung nodule measurements in vivo using semiautomated volumetric measurements. AJR Am J Roentgenol 186: 989–994, 2006.

[9] Gietema HA, Schaefer-Prokop CM, Mali W, Groenewegen G, Prokop M, Pulmonary Nodules: InterscanVariability of Semiautomated Volume Measurements with Multisection CT— Influence of Inspiration Level, Nodule Size, and Segmentation Performance, Radiology: Volume 245: Number 3 December 2007.

[10] Wang Y, van Klaveren RJ, van der Zaag–Loonen HJ, de Bock GH, Gietema HA, Xu DM, Leusveld ALM, de Koning HJ, Scholten ET, Verschakelen J, Prokop M, Oudkerk M, Effect of Nodule Characteristics on Variability of Semiautomated Volume Measurements in Pulmonary Nodules Detected in a Lung Cancer Screening Program, Radiology: Volume 248: Number 2—August 2008.

[11] Zhao, B., James, L.P., Moskowitz, C.S., Guo, P., Ginsberg, M.S., Lefkowitz, R.A., Qin, Y., Riely, G.J., Kris, M.G., Schwartz, L.H. Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. *Radiology* 252: 263–72, 2009.

[12] Hein, P.A., Romano, V.C., Rogalla, P., Klessen, C., Lembcke, A., Dicken, V., Bornemann, L., and Bauknecht, H.C. Linear and volume measurements of pulmonary nodules at different CT dose levels: Intrascan and interscan analysis. Rofo 181: 24–31, 2009.

[13] Mozley PD, Schwartz LH, Bendtsen C, Zhao B, Petrick N, Buckler AJ. Change in lung tumor volume as a biomarker of treatment response: A critical review of the evidence. Annals Oncology; doi:10.1093/annonc/mdq051, March 2010.

[14] Petrou M, Quint LE, Nan B, Baker LH. Pulmonary nodule volumetric measurement variability as a function of CT slice thickness and nodule morphology. Am J Radiol 2007; 188:306-312.

[15] Bogot NR, Kazerooni EA, Kelly AM, Quint LE, Desjardins B, Nan B. Interobserver and intraobserver variability in the assessment of pulmonary nodule size on CT using film and computer display methods. Acad Radiol 2005; 12:948–956.

[16] Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: Implications for assessment of tumor response. J Clin Oncol 2003; 21:2574–2582.

[17] Winer-Muram HT, Jennings SG, Meyer CA, et al. Effect of varying CT section width on volumetric measurement of lung tumors and application of compensatory equations. Radiology 2003; 229:184-194.

[18] Buckler AJ, Mozley PD, Schwartz L, et al. Volumetric CT in lung disease: An example for the qualification of imaging as a biomarker. Acad Radiol 2010; 17:107-115.

[19] AMERICAN COLLEGE OF RADIOLOGY IMAGING NETWORK, ACRIN 6678, FDG-PET/CT as a Predictive Marker of Tumor Response and Patient Outcome: Prospective Validation in Non-small Cell Lung Cancer, August 13, 2010.

[20] Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981;47:207-214.

[21] Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumors: Revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228-247.

[22] McNitt-Gray MF. AAPM/RSNA Physics Tutorial for Residents: Topics in CT. Radiation dose in CT. Radiographics 2002;22:1541-1553.

[23] Xie L, O'Sullivan J, Williamson J, Politte D, Whiting B, TU-FF-A4-02: Impact of Sinogram Modeling Inaccuracies On Image Quality in X-Ray CT Imaging Using the Alternating Minimization Algorithm, Med. Phys. 34, 2571 (2007); doi:10.1118/1.2761438.

[24] Moertel CG, Hanley JA. The effect of measuring error on the results of therapeutic trials in advanced cancer. Cancer 38:388-94, 1976.

[25] Lavin PT, Flowerdew G: Studies in variation associated with the measurement of solid tumors. Cancer 46:1286-1290, 1980.

[26] Eisenhauera EA, Therasseb P, Bogaertsc J, et a. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer 2009; 45: 228-247.

[27] Boll, D.T., Gilkeson, R.C., Fleiter, T.R., Blackham, K.A., Duerk, J.L., and Lewin, J.S. Volumetric assessment of pulmonary nodules with ECG-gated MDCT. AJR Am J Roentgenol 183: 1217–1223, 2004.

[28] Meyer, C.R., Johnson, T.D., McLennan, G., Aberle, D.R., Kazerooni, E.A., Macmahon, H., Mullan, B.F., Yankelevitz, D.F., van Beek, E.J., Armato, S.G., 3rd, McNitt-Gray, M.F., Reeves, A.P., Gur, D., Henschke, C.I., Hoffman, E.A., Bland, P.H., Laderach, G., Pais, R., Qing, D., Piker, C., Guo, J., Starkey, A., Max, D., Croft, B.Y., and Clarke, L.P. Evaluation of lung MDCT nodule annotation across radiologists and methods. Acad Radiol 13: 1254–1265, 2006.

[29] Zhao, B., Schwartz, L.H., Moskowitz, C.S., Ginsberg, M.S., Rizvi, N.A., and Kris, M.G. Lung cancer: computerized quantification of tumor response--initial results. Radiology 241: 892–898, 2006.

[30] Zhao, B., Oxnard, G.R., Moskowitz, C.S., Kris, M.G., Pao, W., Guo, P., Rusch, V.W., Ladanyi, M., Rizvi, N.A., and Schwartz, L.H. A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development. Clin Cancer Res 16: 4647–4653, 2010.

[31] Schwartz, L.H., Curran, S., Trocola, R., Randazzo, J., Ilson, D., Kelsen, D., and Shah, M. Volumetric 3D CT analysis - an early predictor of response to therapy. J Clin Oncol 25: abstr 4576, 2007.

[32] Altorki, N., Lane, M.E., Bauer, T., Lee, P.C., Guarino, M.J., Pass, H., Felip, E., Peylan-Ramu, N., Gurpide, A., Grannis, F.W., Mitchell, J.D., Tachdjian, S., Swann, R.S., Huff, A., Roychowdhury, D.F., Reeves, A., Ottesen, L.H., and Yankelevitz, D.F. Phase II proof-of-concept study of pazopanib monotherapy in treatment-naive patients with stage I/II resectable non-small-cell lung cancer. J Clin Oncol 28: 3131–3137, 2010.

Appendices

Acknowledgements and Attributions

This document is proffered by the Radiological Society of North America (RSNA) Quantitative Imaging Biomarker Alliance (QIBA) Volumetric Computed Tomography (v-CT) Technical Committee. The v-CT technical committee is composed of scientists representing the imaging device manufacturers, image analysis software developers, image analysis laboratories, biopharmaceutical industry, academia, government research organizations, professional societies, and regulatory agencies, among others. All work is classified as pre-competitive. A more detailed description of the v-CT group and its work can be found at the following web link: http://qibawiki.rsna.org/index.php?title=Volumetric_CT.

The Volumetric CT Technical Committee (in alphabetical order):

- Athelogou, M. Definiens AG
- Avila, R. Kitware, Inc.
- Beaumont, H. Median Technologies
- Borradaile, K. Core Lab Partners
- Buckler, A. BBMSC
- Clunie, D. Core Lab Partners
- Cole, P. Imagepace
- Dorfman, G. Weill Cornell Medical College
- Fenimore, C. Nat Inst Standards & Technology
- Ford, R. Princeton Radiology Associates.
- Garg, K. University of Colorado
- Garrett, P. Smith Consulting, LLC
- Gottlieb, R. University of Arizona
- Gustafson, D. Intio, Inc.
- Hayes, W. Bristol Myers Squibb
- Hillman, B. Metrix, Inc.
- Judy, P. Brigham and Women’s Hospital
- Kim, HG. University of California Los Angeles
- Kohl, G. Siemens AG
- Lehner, O. Definiens AG
- Lu, J. Nat Inst Standards & Technology

- McNitt-Gray, M. University California Los Angeles
- Mozley, PD. Merck & Co Inc.
- Mulshine, JL. Rush
- Nicholson, D. Definiens AG
- O'Donnell, K. Toshiba Medical Research Institute - USA
- O'Neal, M. Core Lab Partners
- Petrick, N. US Food and Drug Administration
- Reeves, A. Cornell University
- Richard, S. Duke University
- Rong, Y. Perceptive Informatics, Inc.
- Schwartz, LH. Columbia University
- Saiprasad, G. University of Maryland
- Samei, E. Duke University
- Siegel, E. University of Maryland
- Sullivan, DC. RSNA Science Advisor and Duke University
- Tang, Y. CCS Associates
- Thorn, M. Siemens AG
- Yankellivitz, D. Mt. Sinai School of Medicine
- Yoshida, H. Harvard MGH
- Zhao, B. Columbia University

The Volumetric CT Technical Committee is deeply grateful for the support and technical assistance provided by the staff of the Radiological Society of North America.

Background Information

QIBA

The Quantitative Imaging Biomarker Alliance (QIBA) is an initiative to promote the use of standards to reduce variability and improve performance of quantitative imaging in medicine. QIBA provides a forum for volunteer committees of care providers, medical physicists, imaging innovators in the device and software industry, pharmaceutical companies, and other stakeholders in several clinical and operational domains to reach consensus on standards-based solutions to critical quantification issues. QIBA publishes the specifications they produce (called QIBA profiles), first to gather public comment and then for field test by vendors and users.

QIBA envisions providing a process for developers to test their implementations of QIBA profiles through a compliance mechanism. After a committee determines that a profile has undergone sufficient successful testing and deployment in real-world care settings, it is released for use. Purchasers can specify conformance with appropriate QIBA profiles as a requirement in requests for proposal. Vendors who have successfully implemented QIBA profiles in their products can publish conformance statements (called QIBA Compliance Statements) represented as an appendix called “Model-specific Parameters.” General information about QIBA, including its governance structure, sponsorship, member organizations and work process, is available at http://qibawiki.rsna.org/index.php?title=Main_Page.

CT Volumetry for Cancer Response Assessment

Anatomic imaging using computed tomography (CT) has been historically used to assess tumor burden and to determine tumor response (or progression) to treatment based on uni-dimensional or bi-dimensional measurements. The original WHO response criteria were based on bi-dimensional measurements of the tumor and defined response as a decrease of the sum of the product of the longest perpendicular diameters of measured lesions by at least 50%. The rationale for using a 50% threshold value for definition of response was based on data evaluating the reproducibility of measurements of tumor size by palpation and on planar chest x-rays [24][25]. The more recent RECIST criteria introduced by the National Cancer Institute (NCI) and the European Organisation for Research and Treatment of Cancer (EORTC) standardized imaging techniques for anatomic response assessment by specifying minimum size thresholds for measurable lesions and considered other imaging modalities beyond CT. As well, the RECIST criteria replace longest bi-directional diameters with longest uni-dimensional diameter as the representation of a measured lesion [26]. RECIST defines response as a 30% decrease of the largest diameter of the tumor. For a spherical lesion, this is equivalent to a 50% decrease of the product of two diameters. Current response criteria were designed to ensure a standardized classification of tumor shrinkage after completion of therapy. They have not been developed on the basis of clinical trials correlating tumor shrinkage with patient outcome.

Technological advances in signal processing and the engineering of multi-detector row computed tomography (MDCT) devices have resulted in the ability to acquire high-resolution images rapidly, resulting in volumetric scanning of anatomic regions in a single breath-hold. Volume measurements may be a more sensitive technique for detecting longitudinal changes in tumor masses than reliance on linear tumor diameters as defined by RECIST. Comparative analyses in the context of real clinical trial data have found volume measurements to be more reliable and often more sensitive to longitudinal changes in response than the use of diameters in RECIST. As a result of this increased detection sensitivity and reliability, volume measurements may improve the predictability of clinical outcomes during therapy compared with RECIST. Volume measurements could also benefit patients who need alternative treatments when their disease stops responding to their current regimens [29-32].

The rationale for volumetric approaches to accessing assessing longitudinal changes in tumor burden is multi-factorial. First, most cancers may grow and regress irregularly in three dimensions. Measurements obtained in the transverse plane fail to account for growth or regression in the longitudinal axis, whereas volumetric measurements incorporate changes in all dimensions. Secondly, changes in volume are less subject to either reader error or inter-scan variations. For example, partial response using the RECIST criteria requires a greater than 30% decrease in tumor diameter, which corresponds to greater than 50% reduction in volume of tumor. If one assumes a 21 mm diameter lesion (of 4850 mm3 volume), partial response would result require that the tumor shrink to a in a diameter of less than 158 mm, but which would correspond to a decrease in volume all the way down to 17702145 mm3. The much greater absolute magnitude of volumetric changes is potentially less prone to measurement error than changes in diameter, particularly if the lesions are irregularly shaped or spiculated. As a result of the observed increased sensitivity and reproducibility, volume measurements may be more suited than uni-dimensional measurements to identify early changes in patients undergoing treatment.

Table Summarizing Precision/reproducibility of volumetric measurements from clinical studies reported in the literature

Scan	Reader	# of Readers	# of Patients	# of Nodules	Lesion Size, Mean (range)	Organ System	Volumetry, 95% CI of Measurement Difference	Volumetry, Measurement Difference %	1D Measurement, 95% CI of Measurement Difference	1D, Mean Measurement Difference %	Slice Thickness /Recon Interval, mm	Author, Year
repeat scans	intra-reader	1	20	218	9.85 mm	lung, mets	-21.2 to 23.8%	1.30%			1.0/0.7	Gietama *et al.* 2007 [9]
repeat scans	intra-reader	3	32	32	38 mm (11–93 mm)	lung, NSCLC	-12 to 13.4%	0.70%	-7.3% to 6.2%	-0.60%	1.25/1.25	Zhao *et al.* 2009 [11]
same scan	intra-reader	1	10	50	6.9 mm (2.2–20.5 mm)	lung, mets	-3.9 to 5.7%	0.90%	not reported	not reported	1.25/0.8	Wormanns *et al.* 2004 [4]
same scan	inter-reader	2	10	50	6.9 mm (2.2–20.5 mm)	lung, mets	-5.5 to 6.6%	0.50%	not reported	not reported	1.25/0.8	Wormanns *et al.* 2004 [4]
repeat scans	not specified	not specified	10	151	7.4 (2.2–20.5 mm)	lung, mets	-20.4 to 21.9%	1.50%	not reported	not reported	1.25/0.8	Wormanns *et al.* 2004 [4]
repeat scans	not specified	not specified	10	105	<10 mm	lung, mets	-19.3 to 20.4%	1.70%	not reported	not reported	1.25/0.8	Wormanns *et al.* 2004 [4]
same scan (5 sets, 1 set/phase)	intra-reader ? (consensus by 2 readers), 3 x reading	2	30	73	~1–9 mm [25.3 (0.2–399 mm^3)]	lung, noncalcified nodules	coefficient of variance as large as 34.5% (95% CI not reported)	not reported	not reported	not reported	0.75/0.6	Boll *et al.* 2004 [27]
same scan	inter-reader	2	33	229	10.8 mm (2.8–43.6 mm), median 8.2 mm	lung, primary or mets	-9.4 to 8.0%	0.70%	-31.0 to 27%	-2.00%	1.0/0.8	Hein *et al.* 2009 [12]
same scan	inter-reader, inter-algorithms (6 readers x 3 algorithms)	6	16	23	not reported	lung, nodules	55% (upper limit)	not reported	not reported	not reported	1.25/0.625	Meyer *et al.* 2006 [28]
same scan	intra-reader	2	50	202	3.16–5195 mm^3, median 182.22 mm^3	lung, mets	% not reported	0.15 to 0.22%	% not reported	2.34–3.73% (p<0.05 1D vs 3D)	0.75/0.70	Marten *et al.* 2006 [7]
same scan	inter-reader	2	50	202	3.16–5195 mm^3, median 182.22 mm^3	lung, mets	% not reported	0.22 to 0.29%	% not reported	3.53–3.76% (p<0.05 1D vs 3D)	0.75/0.70	Marten *et al.* 2006 [7]
same scan	inter-reader	2	2239	4225	15–500 mm3 (effective diameter 3.1–	lung, nodules	-13.4 to 14.5%	0.50%	not reported	not reported	1.0/0.7	Wang *et al.* 2008 [10]

Scan	Reader	# of Readers	# of Patients	# of Nodules	Lesion Size, Mean (range)	Organ System	Volumetry, 95% CI of Measurement Difference	Volumetry, Measurement Difference %	1D Measurement, 95% CI of Measurement Difference	1D, Mean Measurement Difference %	Slice Thickness /Recon Interval, mm	Author, Year
					9.8 mm)							
same scan	intra-reader	2	24	52	8.5 mm (<5 to 18 mm)	lung, noncalcified nodules	8.9 % (upper limit)	not reported	not reported	not reported	1.25 or 2.5/not specified	Revel *et al*.[6]
same scan	inter-reader (3 readers x 3 measurements)	3	24	52	8.5 mm (< 18 mm)	lung, noncalcified nodules	6.38 % (upper limit)	not reported	not reported	not reported	1.25 or 2.5/not specified	Revel *et al*. [6]

Abbreviations: 1D = unidimensional; mets = metastasis; CI = confidence interval

Conventions and Definitions

Acquisition vs. Analysis vs. Interpretation: This document organizes acquisition, reconstruction, post-processing, analysis and interpretation as steps in a pipeline that transforms data to information to knowledge. Acquisition, reconstruction and post-processing are considered to address the collection and structuring of new data from the subject. Analysis is primarily considered to be computational steps that transform the data into information, extracting important values. Interpretation is primarily considered to be judgment that transforms the information into knowledge. (The transformation of knowledge into wisdom is beyond the scope of this document.)

Other Definitions:

Image Analysis, Image Review, and/or Read: Procedures and processes that culminate in the generation of imaging outcome measures, such tumor response criteria. Reviews can be performed for eligibility, safety or efficacy. The review paradigm may be context specific and dependent on the specific aims of a trial, the imaging technologies in play, and the stage of drug development, among other parameters.

Image Header: The Image Header is that part of the file or dataset containing the image other than the pixel data itself

Imaging Phantoms: Devices used for periodic testing and standardization of image acquisition. This testing must be site specific and equipment specific and conducted prior to the beginning of a trial (baseline), periodically during the trial and at the end of the trial.

Intra-Rater Variability is the variability in the interpretation of a set of images by the same reader after an adequate period of time inserted to reduce recall bias.

Inter-Rater Variability is the variability in the interpretation of a set of images by the different readers.

A Time Point is a discrete period during the course of a clinical trial when groups of imaging exams or clinical exams are scheduled.

Model-specific Instructions and Parameters

For acquisition modalities, reconstruction software and software analysis tools, profile compliance requires meeting the activity specifications above; e.g. in Sections 2, 3 and 4.

This Appendix provides, as an informative tool, some specific acquisition parameters, reconstruction parameters and analysis software parameters that are expected to be compatible with meeting the profile requirements. Just using these parameters without meeting the requirements specified in the profile is not sufficient to achieve compliance. Conversely, it is possible to use different compatible parameters and still achieve compliance.

These settings were determined to be reasonable by the QIBA CT 1C groundwork study team.

Sites using models listed here are encouraged to consider using these parameters for both simplicity and consistency. Sites using models not listed here may be able to devise their own settings that result in data meeting the requirements.

Table Model-specific Parameters for Acquisition Devices

IMPORTANT NOTE: <u>The presence of a product model/version in the table does not imply it has demonstrated compliance with the QIBA Profile.</u> Refer to the QIBA Compliance Statement for the product.

Acquisition Device	Settings Compatible with Compliance	
GE Discovery HD750 sct3	kVp	120
	Number of Data Channels (N)	64
	Width of Each Data Channel (T, in mm)	0.625
	Gantry Rotation Time in seconds	1
	mA	120
	Pitch	0.984
	Scan FoV	Large Body (500mm)
Philips Brilliance 16 IDT mx8000	kVp	120
	Number of Data Channels (N)	16
	Width of Each Data Channel (T, in mm)	0.75
	Gantry Rotation Time in seconds	0.75
	Effective mAs	50
	Pitch	1.0
	Scan FoV	500
Philips Brilliance 64	kVp	120
	Number of Data Channels (N)	64

Acquisition Device	Settings Compatible with Compliance	
	Width of Each Data Channel (T, in mm)	0.625
	Gantry Rotation Time in seconds	0.5
	Effective mAs	70
	Pitch	0.798
	Scan FoV	500
Siemens Sensation 64	kVp	120
	Collimation (on Operator Console)	64 x 0.6 (Z-flying focal spot)
	Gantry Rotation Time in seconds	0.5
	Effective mAs	100
	Pitch	1.0
	Scan FoV	500
Toshiba Aquilion 64	kVp	120
	Number of Data Channels (N)	64
	Width of Each Data Channel (T, in mm)	0.5
	Gantry Rotation Time in seconds	0.5
	mA	25
	Pitch	.828
	Scan FoV	Medium and Large

Table Model-specific Parameters for Reconstruction Software

IMPORTANT NOTE: The presence of <u>a product model/version in the table does not imply it has demonstrated compliance with the QIBA Profile.</u> Refer to the QIBA Compliance Statement for the product.

Reconstruction Software	Settings Compatible with Compliance	
GE Discovery HD750 sct3	Reconstructed Slice Width, mm	1.25
	Reconstruction Interval	1.0mm
	Display FOV, mm	350
	Recon kernel	STD
Philips Brilliance 16 IDT mx8000	Reconstructed Slice Width, mm	1.00
	Reconstruction Interval	1.0mm (contiguous)
	Display FOV, mm	350

Reconstruction Software	Settings Compatible with Compliance	
	Recon kernel	B
Philips Brilliance 64	Reconstructed Slice Width, mm	1.00
	Reconstruction Interval	1.0mm (contiguous)
	Display FOV, mm	350
	Recon kernel	B
Siemens Sensation 64	Reconstructed Slice Width, mm	1.00
	Reconstruction Interval	1.0mm
	Display FOV, mm	350
	Recon kernel	B30
Toshiba Aquilion 64	Reconstructed Slice Width, mm	1.00
	Reconstruction Interval	1.0mm
	Display FOV, mm	350
	Recon kernel	FC11

Table Model-specific Parameters for Image Analysis Software

IMPORTANT NOTE: The presence of a product model/version in the table does not imply it has demonstrated compliance with the QIBA Profile. Refer to the QIBA Compliance Statement for the product.

Image Analysis Software	Settings Compatible with Compliance	
Siemens LunCARE	a	<settings to achieve…>
	b	<settings to achieve…>
	c	<settings to achieve…>
	d	<settings to achieve…>
GE Lung VCAR	e	<settings to achieve…>
	f	<settings to achieve…>
	g	<settings to achieve…>
	h	<settings to achieve…>
R2 ImageChecker CT Lung	i	<settings to achieve…>
	j	<settings to achieve…>

Image Analysis Software	Settings Compatible with Compliance	
System	k	<settings to achieve…>
	l	<settings to achieve…>
Definiens (name specific product)	m	<settings to achieve…>
	n	<settings to achieve…>
	o	<settings to achieve…>
	p	<settings to achieve…>
Median (name specific product)	q	<settings to achieve…>
	r	<settings to achieve…>
	s	<settings to achieve…>
	t	<settings to achieve…>
Intio (name specific product)	u	<settings to achieve…>
	v	<settings to achieve…>
	w	<settings to achieve…>
	x	<settings to achieve…>