QIBA Profile:

CT Tumor Volume Change (CTV-1)

Version 2.2
8 Aug 2012
Status: (pre)Reviewed

Table of Contents

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.

1	**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. Basically, yes. Claim reworded to be clear and statistically appropriate. The concept of "levels of confidence" has been introduced (See separate documents and process). Claim seems to be appropriate for the "Reviewed" level of confidence. In terms of anatomy, it is recognized that the acquisition protocols and processing will not be appropriate for all types of tumors in all parts of the body, however it is felt that the conspicuity requirements will make it clear to users of the profile which anatomy is not included. E.g. brain tumors will clearly not have sufficient conspicuity. Despite the selection of the acquisition parameters, it is expected that the segmentation algorithms will be able to handle the breadth.
2	**Q. What kind of additional study (if any is needed) would best prove the profile claim?** A. Additional study (as described in the evolving Levels of Confidence document) would provide increased confidence. With this stabilized specification QIBA CT can proceed to such testing.
3	**Q. How do we balance specifying what to accomplish vs how to accomplish it?** 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? A: Have made revisions to text to try to achieve an appropriate balance. The details of compliance testing are still not complete and will require further work in future drafts of the profile.
4	**Q. Should there be a "patient appropriateness" or "subject selection" section?** A. The claim is conditioned upon the lesion being measurable (and criteria are listed) and a section describes characteristics of appropriate (and/or inappropriate) subjects.
5	**Q. Does 4cm/sec "scan speed" preclude too many sites?** A. No. Most 16-slice (and greater) scanners would be able to achieve this (although due to an idiosyncracy of the available scan modes, the total collimation needs to be dropped to 16mm rather than 20mm) Some examples that would meet this include: (a) 16 x 1mm collimation with 0.5 second rotation time and pitch ³ 1.25 OR (b) 16 x 1mm collimation with 0.4 second rotation time and pitch ³ 1 OR (c) 16 x 1.25 mm collimation with 0.5 second rotation time and pitch ³ 1 OR (d) 16 x 1.5mm collimation with 0.5 second rotation time and pitch ³ .833

	Keep in mind that 16 x 0.75 mm collimation would require (i) pitch > 1.67 at 0.5 second rotation time (which breaks the Pitch< 1.5 requirement OR (ii) pitch > 1.33 at 0.4 second rotation time (which is fine) A 4cm/sec threshold is needed since it would likely alleviate potential breath hold issues. Because the reconstructed image thickness allowed here was > 2 mm, all of the above collimation settings would be able to meet both the breath hold requirements as well as the reconstructed image thickness requirements.
6	**Q. What do we mean by noise and how do we measure it?** A. Noise means standard deviation of a region of interest as measured in a homogeneous water phantom. FDA has starting looking at Noise Power Spectrum in light of recent developments in iterative reconstruction and an interest in evaluating what that does to the image quality/characteristics. QIBA should follow what comes out of those discussions, but since FDA is not mandating it and since few systems or sites toda are in a position to measure or make effective use of it, this profile will not mandate it either. It has promise though and would be worth considering for future profile work.
7	**Q. Is 5HU StdDev a reasonable noise value for all organs?** A. No. Will change to 18HU. Not sure where the 5 HU standard deviation came from. The 1C project used a standard deviation of 18HU. At UCLA, our Siemens Sensation 64 will yield a standard deviation of 17 HU for: a. 120kVp, 50 eff. mAs, 1 mm thickness, B30F filter To get this down to 5 HU would require: a. Increasing the eff. mAs to 550, OR b. Increasing the slice thickness to 2 mm AND increasing eff. mAs to 275
8	**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.
9	**Q. Have we worked out the details for how we establish compliance to these specifications?** A. Not completely. 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.
10	**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. For the basis, see the paragraph below the table in Section B.2. It includes reader performance. Allocation of variability across the pipeline (shown in Figure 1) is fraught with difficulty and accounting for reader performance is 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).
11	**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.
12	**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.
13	**Q. What about dose?** A. A discussion has been added in Section 2 to address dose issues.
14	**Q. Are there any IRB questions that should be addressed?** A. The UPICT protocol that will be derived from this Profile will flush out any IRB issues if they exist.
15	**Q. What mechanisms are suggested to achieve consistency with baseline parameters?** A. Basically manual for now. In the future we can consider requiring the parameters be stored in the DICOM image headers or (future) DICOM Protocol Objects, and require systems be able to query/retrieve/import such objects to read prior parameters.

1. Executive Summary

X-ray computed tomography provides an effective imaging technique for assessing treatment response in subjects with cancer. Size quantification is helpful to evaluate tumor changes 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 RECIST have been reported. Investigators have suggested that quantifying whole tumor volumes could solve some of the limitations of diameter measures [1-2] and many studies have explored the value of volumetry [3-12]. This document proposes standardized methods for performing repeatable volume measurements.

This QIBA Profile makes claims about the confidence with which changes in tumor volumes can be measured under a set of defined image acquisition, processing, and analysis conditions, and provides specifications that may be adopted by users and equipment developers to meet targeted levels of clinical performance in identified settings.

The claims are based on several studies of varying scope now underway to provide comparison between the effectiveness of volumetry and uni-dimensional longest diameters as the basis for RECIST in multi-site, multi-scanner-vendor settings.

The intended audiences of this document include:

- Technical staff of software and device manufacturers who create products for this purpose
- Biopharmaceutical companies, oncologists, and clinical trial scientists designing trials with imaging endpoints
- Clinical trialists
- Radiologists, technologists, and administrators at healthcare institutions considering specifications for procuring new CT equipment
- Radiologists, technologists, and physicists designing CT acquisition protocols
- Radiologists and other physicians making quantitative measurements on CT images
- Regulators, oncologists, and others making decisions based on quantitative image measurements

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

2. Clinical Context and Claims

Utilities and Endpoints for Clinical Trials

These specifications are appropriate for quantifying the volumes of malignant tumors and measuring tumor 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

A measured volume change of more than 30% for a tumor provides at least a 95% probability that there is a true volume change; P(true volume change > 0% | measured volume change >30%) > 95%.

This claim holds when the given tumor is measurable (i.e., tumor margins are sufficiently conspicuous and geometrically simple enough to be recognized on all images in both scans), and the longest in-plane diameter of the tumor is 10 mm or greater. Volume change refers to proportional change, where the percentage change is the difference in the two volume measurements divided by the average of the two measurements. By using the average instead of one of the measurements as the denominator, asymmetries in percentage change values are avoided.

Procedures for claiming compliance to the Image Data Acquisition and Image Data Reconstruction activities have been provided (See Section 4). Procedures for claiming compliance to the Image Analysis activity are proposed in draft form and will be revised in the future.

For details on the derivation and implications of the Claim, refer to Appendix B.

While the claim has been informed by an extensive review of the literature, it is currently a consensus claim that has not yet been fully substantiated by studies that strictly conform to the specifications given here. A standard utilized by a sufficient number of studies does not exist to date. 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.

3. Profile Details

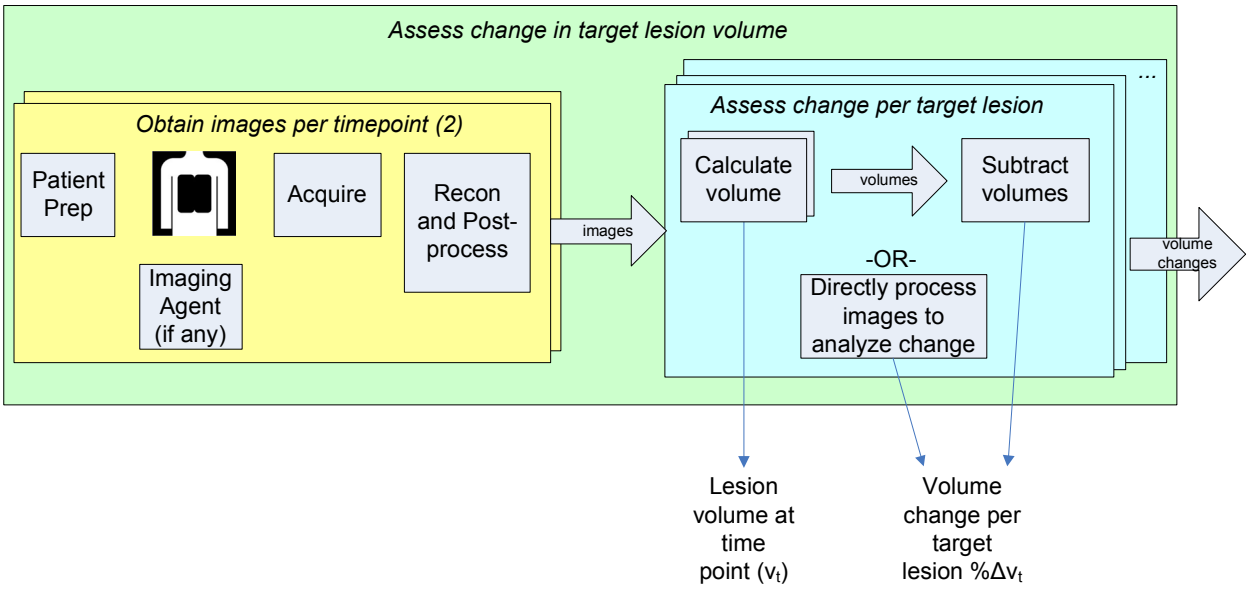
The Profile is documented in terms of “Actors” performing “Activities”.

Equipment, software, staff or sites may claim conformance to this Profile as one or more of the “Actors” in the following table. Compliant Actors shall support the listed Activities by meeting all requirements in the referenced Section. Failing to comply with a “shall” is a protocol deviation. Although deviations invalidate the Profile Claim, such deviations may be reasonable and unavoidable as discussed below.

105 **Table 1: Actors and Required Activities**

Actor	Activity	Section
Acquisition Device	Subject Handling	3.1.
	Image Data Acquisition	3.2.
Technologist	Subject Handling	3.1.
	Image Data Acquisition	3.2.
	Image Data Reconstruction	3.3.
Radiologist	Subject Handling	3.1.
	Image Analysis	3.4.
Reconstruction Software	Image Data Reconstruction	3.3.
Image Analysis Tool	Image Analysis	3.4.

106 The sequencing of the Activities specified in this Profile are shown in Figure 1:



107

108 **Figure 1: CT Tumor Volumetry - Activity Sequence**

109 The method for measuring change in tumor volume may be described as a pipeline. Subjects are prepared
110 for scanning, raw image data is acquired, images are reconstructed and possibly post-processed. Such
111 images are obtained at two (or more) time points. Image analysis assesses the degree of change between
112 two time points for each evaluable target lesion by calculating absolute volume at each time point and
113 subtracting. Volume change is expressed as a percentage (delta volume between the two time points
114 divided by the average of the volume at time point 1 and time point t).

The change may be interpreted according to a variety of different response criteria. These response criteria are beyond the scope of this document. Detection and classification of lesions as target is also beyond the scope of this document.

The Profile does not intend to discourage innovation. The above pipeline provides a reference model. Algorithms which achieve the same result as the reference model but use different methods are permitted, for example by directly measuring the change between two image sets rather than measuring the absolute volumes separately.

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 does not intend to limit how equipment suppliers meet these requirements.

This Profile is "lesion-oriented". The Profile requires that images of a given tumor be acquired and processed the same way each time. It does not require that images of tumor A be acquired and processed the same way as images of tumor B; for example, tumors in different anatomic regions may be imaged or processed differently, or some tumors might be examined at one contrast phase and other tumors at another phase.

The requirements in this Profile do not codify a Standard of Care; they only provide guidance intended to achieve the stated Claim. Although deviating from the specifications in this Profile may invalidate the Profile Claims, the radiologist or supervising physician is expected to do so when required by the best interest of the patient or research subject. How study sponsors and others decide to handle deviations for their own purposes is entirely up to them.

Since much of this Profile emphasizes performing subsequent scans consistent with the baseline scan of the subject, the parameter values chosen for the baseline scan are particularly significant and should be considered carefully.

In some scenarios, the "baseline" might be defined as a reference point that is not necessarily the first scan of the patient.

3.1. Subject Handling

This Profile will refer primarily to "subjects", keeping in mind that the requirements and recommendations apply to patients in general, and subjects are often patients too.

3.1.1 Timing Relative to Index Intervention Activity

When the Profile is being used in the context of a clinical trial, refer to relevant clinical trial protocol for further guidance or requirements on timing relative to index intervention activity.

3.1.2 Timing Relative to Confounding Activities

This document does not presume any other timing relative to other activities.

Fasting prior to a contemporaneous FDG PET scan or the administration of oral contrast for abdominal CT is not expected to have any adverse impact on this Profile.

3.1.3 Contrast Preparation and Administration

3.1.3.1 DISCUSSION

Contrast characteristics influence the appearance, conspicuity, and quantification of tumor volumes. Non-contrast CT may not permit an accurate characterization of the malignant visceral/nodal/soft-tissue lesions and assessment of tumor boundaries. Therefore, consistent use of intravenous contrast is required to meet the claims of this Profile.

However, the use of contrast material (intravenous or oral) may be not be medically indicated in defined clinical settings or may be contra-indicated for some subjects. Radiologists and supervising physicians may omit intravenous contrast or vary administration parameters when required by the best interest of patients or research subjects, in which case lesions may still be measured but the measurements will not be subject to the Profile claims.

The following specifications are minimum requirements to meet Profile claims. Ideally, intravenous contrast type, volume, injection rate, use or lack of a "saline chase," and time between contrast administration and image acquisition should be identical for all time points, and the use of oral contrast material should be consistent for all abdominal imaging timepoints.

Recording the use and type of contrast, actual dose administered, injection rate, and delay in the image header by the Acquisition Device is recommended. This may be by automatic interface with contrast administration devices in combination with text entry fields filled in by the Technologist. Alternatively, the technologist may enter this information manually on a form that is scanned and included with the image data as a DICOM Secondary Capture image.

3.1.3.2 SPECIFICATION

Parameter	Specification
Use of intravenous or oral contrast	The Radiologist shall determine if the contrast protocol is appropriate for the subject. The Technologist shall use intravenous contrast parameters consistent with baseline. Specifically, the total amount of contrast administered (grams of iodine) shall not vary by more than 25% between scans; contrast injection rate shall be at least 2ml/sec and shall not vary by more than 2ml/sec for arterial phase imaging, and images to be compared are to be obtained at the same phase (i.e. arterial, venous, or delayed).

3.1.4 Subject Positioning

3.1.4.1 DISCUSSION

Consistent positioning avoids unnecessary changes 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 arms, the anterior-to-posterior curvature of their spines as

determined by pillows under their backs or knees, the lateral straightness of their spines. Prone positioning is not recommended. 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.

When the patient is supine, the use of positioning wedges under the knees and head is recommended so that the lumbar lordosis is straightened and the scapulae are both in contact with the table. However, the exact size, shape, etc. of the pillows is not expected to significantly impact the Profile Claim. It is expected that clinical trial documentation or local clinical practice will specify their preferred patient positioning.

Recording the Subject Positioning and Table Heights in the image header is helpful for auditing and repeating baseline characteristics.

Consistent centering of the patient avoids unnecessary variation in the behavior of dose modulation algorithms during scan.

3.1.4.2 SPECIFICATION

Parameter	Specification
Subject Positioning	The Technologist shall position the subject consistent with baseline. If baseline positioning is unknown, position the subject Supine if possible, with devices such as positioning wedges placed as described above.
Table Height & Centering	The Technologist shall adjust the table height for the mid-axillary plane to pass through the isocenter. The Technologist shall position the patient such that the “sagittal laser line” lies along the sternum (e.g. from the suprasternal notch to the xiphoid process).

3.1.5 Instructions to Subject During Acquisition

3.1.5.1 DISCUSSION

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

Since some motion may occur due to diaphragmatic relaxation in the first few seconds following full inspiration, a proper breath hold will include instructions like "Lie still, breathe in fully, hold your breath, and relax”, allowing 5 seconds after achieving full inspiration before initiating the acquisition.

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 tumors are located.

3.1.5.2 SPECIFICATION

Parameter	Specification

Parameter	Specification
Breath hold	The Technologist shall instruct the subject in proper breath-hold and start image acquisition shortly after full inspiration, taking into account the lag time between full inspiration and diaphragmatic relaxation. The Technologist shall ensure that for each tumor the breath hold state is consistent with baseline.
Image Header	The Technologist shall record factors that adversely influence subject positioning or limit their ability to cooperate (e.g., breath hold, remaining motionless, agitation in subjects with decreased levels of consciousness, subjects with chronic pain syndromes, etc.). The Acquisition Device shall provide corresponding data entry fields.

3.1.6 Timing/Triggers

3.1.6.1 Discussion

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

3.1.6.2 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 consistent with baseline.
Image Header	The Acquisition Device shall record actual Timing and Triggers in the image header.

3.2. Image Data Acquisition

3.2.1 Discussion

CT scans for tumor volumetric analysis can be performed on any equipment that complies with the specifications set out in this Profile. However, we strongly encourage performing all CT scans for an individual subject on the same platform (manufacturer, model and version) which we expect will further reduce variation.

Many scan parameters can have direct or indirect effects on identifying, segmenting and measuring lesions. To reduce this potential source of variance, all efforts should be made to have as many of the scan parameters as possible consistent with the baseline.

Consistency with the baseline implies a need for a method to record and communicate the baseline settings and make that information available at the time and place that subsequent scans are performed. Although it is conceivable that the scanner could retrieve prior/baseline images and extract acquisition parameters to encourage consistency, such interoperability mechanisms are not defined or mandated here and cannot be depended on to be present or used. Similarly, managing and forwarding the data files when multiple sites

are involved may exceed the practical capabilities of the participating sites. Sites should be prepared to use manual methods instead.

The goal of parameter consistency is to achieve consistent performance. Parameter consistency when using the same scanner make/model generally means using the same values. Parameter consistency when the baseline was acquired on a *different* make/model may require some "interpretation" to achieve consistent performance since the same values may produce different behavior on different models. The parameter sets in Appendix D may be helpful in this task.

The approach of the specifications here, and in the reconstruction section, is to focus as much as possible on the characteristics of the resulting dataset, rather than one particular technique for achieving those characteristics. This is intended to allow as much flexibility as possible for product innovation and reasonable adjustments for patient size (such as increasing acquisition mAs and reconstruction DFOV for larger patients), while reaching the performance targets. Again, the technique parameter sets in Appendix D may be helpful for those looking for more guidance.

The purpose of the minimum scan speed requirement is to permit acquisition of an anatomic region in a single breath-hold, thereby preventing respiratory motion artifacts or anatomic gaps between breath-holds. This requirement is applicable to scanning of the chest and upper abdomen, the regions subject to these artifacts, and is not required for imaging of the head, neck, pelvis, spine, or extremities.

Coverage of additional required anatomic regions (e.g. to monitor for metastases in areas of likely disease) depends on the requirements of the clinical trial or local clinical practice. In baseline scans, the tumor locations are unknown and may result in a tumor not being fully within a single breath-hold, making it "unmeasurable" in the sense of this Profile.

Pitch is chosen so as to allow completion of the scan in a single breath hold.

For subjects needing two or more breath-holds to fully cover an anatomic region, different tumors may be acquired on different breath-holds. It is still necessary that each tumor be fully included in images acquired within a single breath-hold to avoid discontinuities or gaps that would affect the measurement.

Scan Plane (transaxial is preferred) may differ between subjects due to the need to position for physical deformities or external hardware. For an individual subject, a consistent scan plane will reduce unnecessary differences in the appearance of the tumor.

Total Collimation Width (defined as the total nominal beam width, NxT, for example 64x1.25mm) is often not directly visible in the scanner interface. Manufacturer reference materials typically explain how to determine this for a particular scanner make, model and operating mode. Wider collimation widths can increase coverage and shorten acquisition, but can introduce cone beam artifacts which may degrade image quality. Imaging protocols will seek to strike a balance to preserve image quality while providing sufficient coverage to keep acquisition times short.

Nominal Tomographic Section Thickness (T), the term preferred by the IEC, is sometimes also called the Single Collimation Width. It affects the spatial resolution along the subject z-axis.

Smaller voxels are preferable to reduce partial volume effects and provide higher accuracy due to higher spatial resolution. The resolution/voxel size that reaches the analysis software is affected by both acquisition parameters and reconstruction parameters.

X-ray CT uses ionizing radiation. Exposure to radiation can pose risks; however 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.

Anatomic Coverage recording by the Acquisition Device may or may not require the attention of the Technologist.

The acquisition parameter constraints here have been selected with scans of the chest, abdomen and pelvis in mind.

3.2.2 SPECIFICATION

The Acquisition Device shall be capable of performing scans with all the parameters set as described in the following table. The Technologist shall set up the scan to achieve the requirements in the following table.

Parameter	Specification	DICOM Tag
Scan Duration for Thorax	Achieve a table speed of at least 4cm per second, if table motion is necessary to cover the required anatomy.	Table Speed (0018,9309)
Anatomic Coverage	Tumors to be measured and additional required anatomic regions shall be fully covered. If multiple breath-holds are required, the technologist shall obtain image sets with sufficient overlap to avoid gaps within the required anatomic region(s), and shall ensure that each tumor lies wholly within a single breath-hold.	Anatomic Region Sequence (0008,2218)
Scan Plane (Image Orientation)	Consistent with baseline.	Gantry/Detector Tilt (0018,1120)
Total Collimation Width	Greater than or equal to 16mm.	Total Collimation Width (0018,9307)
IEC Pitch	Less than 1.5.	Spiral Pitch Factor (0018,9311)
Tube Potential (kVp)	Consistent with baseline (i.e. the same kVp setting if available, otherwise as similar as possible).	KVP (0018,0060)
Nominal Tomographic Section Thickness (T)	Less than or equal to 1.5mm.	Single Collimation Width (0018,9306)
Acquisition Field of View (FOV)	Consistent with baseline.	

Parameter	Specification	DICOM Tag
Image Header	The Acquisition Device shall record actual Field of View, Scan Duration, Scan Plane, Total Collimation Width, Single Collimation Width, Scan Pitch, Tube Potential, Tube Current, Rotation Time, Exposure and Slice Width in the DICOM image header.	

3.3. Image Data Reconstruction

3.3.1 Discussion

Image reconstruction is modeled as a separate Activity in the QIBA Profile. Although it is closely related to image acquisition, and is usually performed on the Acquisition Device, reconstruction may be performed, or re-performed, separate from the acquisition. Many reconstruction parameters will be influenced or constrained by related acquisition parameters. This specification is the result of discussions to allow a degree of separation in their consideration without suggesting they are totally independent.

Many reconstruction parameters can have direct or indirect effects on identifying, segmenting and measuring lesions. To reduce this potential source of variance, all efforts should be made to have as many of the parameters as possible consistent with the baseline.

Consistency with the baseline implies a need for a method to record and communicate the baseline settings and make that information available at the time and place that subsequent reconstructions are performed. Although it is conceivable that the scanner could retrieve prior/baseline images and extract reconstruction parameters to encourage consistency, such interoperability mechanisms are not defined or mandated here and cannot be depended on to be present or used. Similarly, managing and forwarding the data files when multiple sites are involved may exceed the practical capabilities of the participating sites. Sites should be prepared to use manual methods instead.

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 which may degrade segmentation and quantification of tumors. 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. Maximum spatial resolution is mostly determined by the scanner geometry (which is not usually under user control) and the reconstruction kernel (over which the user has some choice). Resolution 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)”. If a followup scan has a significantly different resolution than the baseline, it is likely that the exposure characteristics will change which can affect repeatability. The impact of partial volume effects can also change, so reasonable consistency of resolution within a given patient is desirable.

Noise Metrics quantify the magnitude of the random variation in reconstructed CT numbers. Increased levels of noise can make it difficult to identify the boundary of tumors by humans and automated algorithms.

Some properties of the noise can be characterized by the standard deviation of reconstructed CT numbers over a uniform region in phantom. Voxel Noise (pixel standard deviation in a region of interest) can be reduced by reconstructing images with greater thickness for a given mAs. A constant value for the noise metric might be achieved by increasing mAs for thinner reconstructed images and reducing mAs for thicker reconstructed images. The use of a standard deviation metric has limitations since it can vary with different reconstruction kernels, which will also impact the spatial resolution. While the Noise-Power Spectrum would be a more comprehensive metric, it is not practical to calculate (and interpret) at this time. Therefore, the standard deviation metric is the preferred measure for Voxel Noise. It is not expected that the Voxel Noise be measured for each subject scan, but rather the Acquisition Device and Reconstruction Software be qualified for the expected acquisition and reconstruction parameters.

Reconstruction Field of View affects reconstructed pixel size because the fixed image matrix size of most reconstruction algorithms is 512x512. 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 spatial resolution in each dimension. The spatial resolution of the reconstructed image depends on a number of additional factors including a strong dependence on the reconstruction kernel.

Reconstruction Interval (a.k.a. Slice spacing) that results in discontiguous data is unacceptable as it 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 multi-detector 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.

Slice thickness is "nominal" since the thickness is not technically the same at the middle and at the edges.

Reconstruction Kernel Characteristics influence the texture and the appearance of tumors in the reconstructed images, which may influence measurements. 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 characteristics of different tissues (e.g. lung) may call for the use of different kernels, and implementers are encouraged to use kernels suitable for the anatomic region and tissue imaged. The use of multiple kernels in a single study is not prohibited by the specification below, but any given tumor must be measured on images reconstructed using consistent kernels at each time point.

The use of iterative reconstruction also may influence the texture and the appearance of tumors in the reconstructed images, which may influence measurements. Therefore the effects of iterative

reconstruction on quantitative accuracy and reproducibility are not fully understood as of this writing of this Profile version so it is not currently allowed within the Profile Claim.

The stability of HU between time points and its effect on volume measurements is not fully understood as of the writing of this version of the Profile.

3.3.2 SPECIFICATION

The Reconstruction Software shall be capable of producing images that meet the following specifications. The Technologist shall set up or configure the reconstruction to achieve the requirements in the following table.

Parameter	Specification
In-plane Spatial Resolution	Greater than or equal to 6 lp/cm and consistent with baseline (i.e. within 1 lp/cm).
Voxel Noise	Standard deviation of < 18HU measured near the center of a 20cm water phantom.
Reconstruction Field of View	Spanning at least the full extent of the thoracic and abdominal cavity, but not significantly greater than required to show the entire body and consistent with baseline.
Slice Thickness	Less than or equal to 2.5 mm and consistent with baseline (i.e. within 0.5mm).
Reconstruction Interval	Less than or equal to 2.5 mm and consistent with baseline.
Reconstruction Overlap	Greater than or equal to 0 (i.e. no gap, and may have some overlap) and consistent with baseline.
Reconstruction Algorithm Type	Filtered Back-Projection
Reconstruction Kernel Characteristics	Consistent with baseline (i.e. the same kernel if available, otherwise the kernel most closely matching the kernel response of the baseline).
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.

3.4. Image Analysis

3.4.1 DISCUSSION

This Profile characterizes each designated tumor by its volume change relative to prior image sets.

This is typically done by determining the boundary of the tumor (referred to as segmentation), computing the volume of the segmented tumor and calculating the difference of the tumor volume in the current scan and in the baseline scan.

Volume Calculation values from a segmentation may or may not correspond to the total of all the segmented voxels. The algorithm may consider partial volumes, do surface smoothing, tumor or organ modeling, or interpolation of user sculpting of the volume. The algorithm may also pre-process the images prior to segmentation.

Segmentation may be performed automatically by a software algorithm, manually by a human observer, or semi-automatically by an algorithm with human guidance/intervention, for example to identify a starting seed point, stroke, or region, or to edit boundaries.

If a human observer participates in the segmentation, either by determining while looking at the images the proper settings for an automated process, or by manually editing boundaries, the settings for conversion of density into display levels (window and level) should either be fixed during the segmentation process or documented so that observers can apply consistent display settings at future scans (or a different observer for the same scan, if multiple readers will read each scan, as for a clinical trial).

Tumor Volume Change Variability, which is the focus of the Profile Claim, is a key performance parameter for this biomarker. The 30% target is a conservative threshold of measurement variation (the 30% change in the claim is the outside of 95% confidence interval of 15% of measurement variability when sample size is 40 or more). Based on a survey of clinical studies (See Appendix B.2) the 30% target is considered to be reasonable and achievable. In Table B.1, the range between the minimum and maximum values in the 95% CI of the measurement difference column is mostly within +/- 15%. Considering a large study from Wang et al using 2239 patients [15], the 95% confidence interval ranged [-13.4%, 14.5%].

Methods that calculate volume changes directly without calculating volumes at individual time points are acceptable so long as the results are compliant with the specifications set out by this Profile.

The Image Analysis Tool should be prepared to process both the current data and previous data at the same time and support matching up the appearance of each tumor in both data sets in order to derive volume change values. Although it is conceivable that they could be processed separately and the results of prior processing could be imported and a method of automated tagging and matching of the tumors could be implemented, such interoperability mechanisms are not defined or mandated here and cannot be depended on to be present or used.

Storing segmentations and measurement results that can be loaded by an Image Analysis Tool analyzing data collected at a later date is certainly a useful practice as it can save time and cost. For this to happen reliably, the stored format must be compatible and the data must be stored and conveyed. Although DICOM Segmentation objects are appropriate to store tumor segmentations, and DICOM SR objects are appropriate to store measurement results, these standards are not yet widely enough deployed to make support for them mandatory in this Profile. Similarly, conveying the segmentations and measurements from baseline (and other time points prior to the current exam) is not done consistently enough to mandate that it happen and to require their import into the Image Analysis Tool. Managing and forwarding the data files may exceed the practical capabilities of the participating sites.

Image analysis can be performed on any equipment that complies with the specifications set out in this Profile. However, we strongly encourage performing all analysis for an individual subject on the same platform (manufacturer, model and version) which we expect will further reduce variation.

Medical Devices such as the Image Analysis Tool are typically made up of multiple components (the hardware, the operating system, the application software, and various function libraries within those). Changes in any of the components can affect the behavior of the device. In this specification, the "device version" should reflect the total set of components and any changes to components should result in a change in the recorded device version. This device version may thus be different than the product release version that appears in vendor documentation.

For analysis methods that involve an operator (e.g. to draw or edit boundaries, set seed points or adjust parameters), the operator is effectively a component of the system, with an impact on the reproducibility of the measurements, and it is important to record the operator's identify as well. Fully automated analysis software removes that source of variation; although even then, since a human is generally responsible for the final results, they retain the power to approve or reject measurements so their identity should be recorded.

The Tumor Volume Change performance specification below includes the operator performance and is intended to be evaluated based on a typical operator (i.e. without extraordinary training or ability). This should be kept in mind by vendors measuring the performance of their tools and sites validating the performance of their installation. Although the performance of some methods may depend on the judgment and skill of the operator, it is beyond this Profile to specify the qualifications or experience of the operator.

Determination of which tumors should be measured is out of scope for this Profile. Such determination may be specified within a protocol or specified by formal response criteria standards, or may be determined by clinical requirements. Tumors to be measured may be designated by the oncologist or clinical investigator, by a radiologist at a clinical site, by a reader at a central reading facility, or they may be designated automatically by a software analysis tool.

3.4.2 SPECIFICATION

Parameter	Specification
Common Tumor Selection	The Image Analysis Tool shall allow all tumors selected for volume measurement to be unambiguously labeled, so that all readers can assess the same tumors.
Multiple Tumors	The Image Analysis Tool shall allow multiple tumors to be measured, and each measured tumor to be associated with a human-readable identifier that can be used for correlation across time points.
Tumor Volume Change Variability	*The following two specifications are essentially the same, with the first applying to the provider of the tool and the second applying to the site where the tool is used.* The Image Analysis Tool shall demonstrate the ability to measure tumor volume change (according to Figure 1) on data that meets the criteria of the preceding activities with a 95% confidence interval around the measured change of no greater than +/- 30%. The Radiologist (if operator interaction is required by the Image Analysis Tool to perform measurements) shall demonstrate the ability to measure tumor volume change (according to

Parameter	Specification
	Figure 1) on data that meets the criteria of the preceding activities with a 95% confidence interval around the measured change of no greater than +/- 30%.
Result Verification	The Radiologist shall review/approve the measurement results as needed.
Recording	The Image Analysis Tool shall record the percentage volume change relative to baseline for each tumor, the device version and the actual model-specific Analysis Software set-up and configuration parameters utilized. The Image Analysis Tool shall be capable of recording the tumor segmentation as a DICOM Segmentation. The Image Analysis Tool shall record the identity of each individual making and/or approving a tumor measurement using the software.

423
424

4. Compliance

To comply with this Profile, participating staff and equipment ("Actors") shall support each of the activities assigned to them in Table 1.

For each activity, the compliance requirements (sometimes referred to as the "shall language") for each Actor are documented in Section 3.

Although most of the requirements described in Section 3 are feature-oriented and compliance can be assessed by direct observation, some of the requirements are performance-oriented. The following sub-sections elaborate on the meaning of performance-oriented requirements and how they are intended to be correctly assessed.

Formal claims of compliance by the organization responsible for an Actor shall be in the form of a published QIBA Conformance Statement. Vendors publishing a QIBA Conformance Statement shall provide a set of "Model-specific Parameters" (as shown in Appendix D) describing how their product was configured to achieve compliance. Vendors shall also provide access or describe the characteristics of the test set used for compliance testing.

4.1. Performance Assessment: Tumor Volume Change Variability

Note: The procedure in this section is currently only a proposal.
A more detailed procedure and pointers to valid test datasets will be provided in the future.
Until then, there is no approved way to claim conformance to this performance requirement.

Tumor Volume Change Variability performance can be assessed with the following procedure:

- Obtain a designated test image set (see 4.1.1).
- Determine the volume change for designated tumors (see 4.1.2).
- Calculate descriptive statistics (see 4.1.3).
- Compare against the Tumor Volume Change Variability performance level specified in 3.4.2.

This procedure can be used by a vendor or an imaging site to evaluate the performance of an Image Analysis Tool (in automatic mode, or with a typical operator), or the combined performance of an Image Analysis Tool together with a particular Radiologist to determine if they are in compliance with the Tumor Volume Change Variability performance requirement in Section 3.4.2.

4.1.1 Test Image Set

The test image set consists of multiple target tumors in multiple body parts in multiple subjects (human or phantom). The body parts are representative of the stated scope of the Profile (i.e. includes lung nodules as well as metastases such as mediastinal, liver, adrenal, neck, axillary, mesenteric, retroperitioneal, pelvic, etc. described in Appendix B.3).

The target tumors are selected to be measureable (i.e. larger than 10mm diameter, geometrically simple

and with sufficiently conspicuous margins) and have a range of volumes, shapes and types to be representative of the scope of the Profile.

The test image set includes at least N target tumors. Each target tumor has at least T time points. The tumors span at least B body parts.

The test image set has been acquired according to the requirements of this Profile (e.g. patient handling, acquisition protocol, reconstruction).

Discussion:
We have many test image cases where the true change is known to be 0% ("Coffee break").
We have many test image cases where the true change is unknown (although change is clearly present).
Are we missing data to show both sensitivity and specificity?
What exactly is our goal with this performance assessment?
Consider a multi- step assessment?

1) Assess (change?) sensitivity (in terms of inherent measurement variation) using "No change" data
2) Assess (volume?) bias using data with a known volume (phantom?)
3) Assess change performance against consensus values (rather than measured/known truth?)

Tumor segmentation performance can be affected by the accuracy or variations in the seed point or axis provided. Consider preparing the test set with test "inputs" (either with a "click here" dot on the image, or some method for feeding coordinates to the application).
Ideally we want fully realistic images (not phantom) but with known truth for tumor volume change. Would it be possible to digitally insert tumors into real acquired human images?

What is the best way to go about assembling and hosting these datasets? Such a public dataset is not currently known to exist.

4.1.2 Determine Volume Change

Determine the measured proportional percentage volume change for each designated tumor in each image multiple times by multiple readers.

Discussion:
Should the (minimum) number of readers and the (minimum) number of repeats for each reader (for each tumor?) be prescribed in the procedure?
Will those numbers be different for fully automated measurements (which are presumably more consistent among repeats on the same data but are generally cheap to run more repeats.)?

Consider whether the procedure should allow a small number of segmentation or volume change results to be set aside prior to calculation of the descriptive statistics to avoid a couple unusual cases from distorting the summary statistics. Such "failures" could still be reported individually in the results.
Would such "blow ups" be easily distinguished by the algorithm or operator? Dan Barboriak has done work on related issues.

4.1.3 Calculate Descriptive Statistics

Calculate descriptive statistics that represent the joint-distribution of true proportional percentage volume change and measured proportional percentage volume change.

Discussion:
The performance score statistics should not be a simple total of all the lesion change vales, but rather we should quote performance on individual lesions over a specified number of repeats for a specified number of lesions.

Given the volume measure at Time1 and Time2, consider both the variance and the correlation between the two measurements (i.e. the variance of the individual measurements and also
*(sigma of the delta)**2 = 2 (1-rho) sigma**2*
It is expected that correlation across visits will be dominated by using a different device?

Consider calculating and expressing in terms of the confidence that a change of size X is really more than Y. ie. in the P(A|B)>C can we fix or "vectorize" any of the three variables? Note that the target zones for change confidence might be different for clinical trials vs patient management. Does this point us toward two claims? Or maybe a claim in the form of a vector of values or a curve?

Alternatively, consider (as suggested by TSB in comment #164) evaluating performance relative to a specified (e.g. expert consensus derived) "truth" value.

Keep in mind that we need to maintain consistency between our claim and our performance measures (e.g. focus on repeatability vs. accuracy).

It is important to characterize individual volume measurement performance since that value is an input to a variety of models (and would be useful for patient enrichment in trials). So, for example:
For each tumor(t)
Average the (r) measurements of t
Enumerate the number of measurements N(t) that are within 30% of the average
N=Sum N(t)
*If N >= 95% of t*r then the 95% confidence performance specification has been met.*

It might be useful to explore the Visual Analog Scale (VAS Score) as a categorization tool for the target tumors and set different variance or performance targets for each category, or consider weighting the errors based on the VAS Score.

4.2. Performance Assessment: Image Acquisition Site

Note: The procedure in this section is currently only a proposal.
A more detailed procedure and pointers to valid test datasets will be provided in the future.
Until then, there is no approved way to claim conformance to this performance requirement.

Site performance can be assessed with the following procedure:

- Validate image acquisition (see 4.2.1).
- Generate a test image set (see 4.2.2).
- Assess Tumor Volume Change Variability (see 4.1.2, 4.1.3 above).
- Compare against the Tumor Volume Change Variability performance level specified in 3.4.2.

This procedure can be used by an imaging site to evaluate the performance of each of the Actors and Activities in use. In principle, the final result represents an assessment of the combined performance of all the Actors and Activities at the site.

The procedure presumes that the Actors being used by the site are capable of meeting the requirements described in Section 3 of this document; however it is not a pre-requisite that those Actors have published QIBA Conformance Statements (although that would be both useful and encouraging).

Discussion:
Duke is working on a "platform" that includes a phantom and an analysis tool that may inform the future contents of this section.

Sites that carry out this procedure should really record the parameters they used and document them in something similar to a Conformance Statement. This would be a useful QA record and could be submitted to clinical trials looking for QIBA compliant test sites.

Are there other criteria that should be worked into this procedure?
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 clinical trial might specify "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."

585 4.2.1 ACQUISITION VALIDATION

586 Review patient handling procedures for compliance with Section 3.1

587 Establish acquisition protocols and reconstruction settings on the Acquisition Device compliant with Section
588 3.2 and Section 3.3. If a QIBA Conformance Statement is available from the Acquisition Device vendor, it
589 may provide parameters useful for this step.

590 Acquire images of a 20cm water phantom, reconstruct and confirm performance requirements in Section
591 3.3.2 are met.

592 *Discussion:*
593 *UCLA may have more detailed and more complete procedures to recommend for this section.*

594 4.2.2 TEST IMAGE SET

595 Locally acquire a test image set using the protocols established and tested in Section 4.2.1.
596
597 The test image set should conform to the characteristics described in Section 4.1.1.
598

599 *Discussion:*
600 *It is highly likely that due to practical constraints the test image set prepared at an individual site would be*
601 *much less comprehensive than the test image sets prepared by QIBA. Further consideration of what a more*
602 *limited but still useful test image set would look like.*
603
604

605

References

1. Moertel, C.G. and J.A. Hanley, *The effect of measuring error on the results of therapeutic trials in advanced cancer.* Cancer, 1976. **38**(1): p. 388-94.
2. Lavin, P.T. and G. Flowerdew, *Studies in variation associated with the measurement of solid tumors.* Cancer, 1980. **46**(5): p. 1286-90.
3. Eisenhauer, E.A., et al., *New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1).* Eur J Cancer, 2009. **45**(2): p. 228-47.
4. Zhao, B., et al., *Lung cancer: computerized quantification of tumor response--initial results.* Radiology, 2006. **241**(3): p. 892-8.
5. Zhao, B., et al., *A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development.* Clin Cancer Res, 2010. **16**(18): p. 4647-53.
6. Schwartz, L.H., et al., *Volumetric 3D CT analysis - an early predictor of response to therapy.* J Clin Oncol, 2007. **25**(18s): p. abstr 4576.
7. Altorki, N., et al., *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, 2010. **28**(19): p. 3131-7.
8. Gietema, H.A., et al., *Pulmonary nodules: Interscan variability of semiautomated volume measurements with multisection CT-- influence of inspiration level, nodule size, and segmentation performance.* Radiology, 2007. **245**(3): p. 888-94.
9. Zhao, B., L.H. Schwartz, and S.M. Larson, *Imaging surrogates of tumor response to therapy: anatomic and functional biomarkers.* J Nucl Med, 2009. **50**(2): p. 239-49.
10. Wormanns, D., et al., *Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility.* Eur Radiol, 2004. **14**(1): p. 86-92.
11. Boll, D.T., et al., *Volumetric assessment of pulmonary nodules with ECG-gated MDCT.* AJR Am J Roentgenol, 2004. **183**(5): p. 1217-23.
12. Hein, P.A., et al., *Linear and volume measurements of pulmonary nodules at different CT dose levels - intrascan and interscan analysis.* Rofo, 2009. **181**(1): p. 24-31.
13. Meyer, C.R., et al., *Evaluation of lung MDCT nodule annotation across radiologists and methods.* Acad Radiol, 2006. **13**(10): p. 1254-65.
14. Marten, K., et al., *Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria.* Eur Radiol, 2006. **16**(4): p. 781-90.
15. Wang, Y., et al., *Effect of nodule characteristics on variability of semiautomated volume measurements in pulmonary nodules detected in a lung cancer screening program.* Radiology, 2008. **248**(2): p. 625-31.
16. Revel, M.P., et al., *Pulmonary nodules: preliminary experience with three-dimensional evaluation.* Radiology, 2004. **231**(2): p. 459-66.
17. Sohns, C., et al., *Measurement of response of pulmonal tumors in 64-slice MDCT.* Acta Radiol, 2010. **51**(5): p. 512-21.
18. Werner-Wasik, M., et al., *Assessment of lung cancer response after nonoperative therapy: tumor diameter, bidimensional product, and volume. A serial CT scan-based study.* Int J Radiat Oncol Biol Phys, 2001. **51**(1): p. 56-61.
19. Tran, L.N., et al., *Comparison of treatment response classifications between unidimensional, bidimensional, and volumetric measurements of metastatic lung lesions on chest computed*

tomography. Acad Radiol, 2004. **11**(12): p. 1355-60.
20. Jennings, P., et al., *Inter-laboratory comparison of human renal proximal tubule (HK-2) transcriptome alterations due to Cyclosporine A exposure and medium exhaustion.* Toxicol In Vitro, 2009. **23**(3): p. 486-99.
21. Prasad, S.R., et al., *CT tumor measurement for therapeutic response assessment: comparison of unidimensional, bidimensional, and volumetric techniques initial observations.* Radiology, 2002. **225**(2): p. 416-9.
22. Rohde, S., et al., *Reliability of CT-based tumor volumetry after intraarterial chemotherapy in patients with small carcinoma of the oral cavity and the oropharynx.* Neuroradiology, 2006. **48**(6): p. 415-21.
23. Lee, S.M., et al., *Usefulness of CT volumetry for primary gastric lesions in predicting pathologic response to neoadjuvant chemotherapy in advanced gastric cancer.* Abdom Imaging, 2009. **34**(4): p. 430-40.
24. Beer, A.J., et al., *Adenocarcinomas of esophagogastric junction: multi-detector row CT to evaluate early response to neoadjuvant chemotherapy.* Radiology, 2006. **239**(2): p. 472-80.
25. Griffith, J.F., et al., *Assessing chemotherapy response of squamous cell oesophageal carcinoma with spiral CT.* Br J Radiol, 1999. **72**(859): p. 678-84.
26. Benz, M.R., et al., *Combined assessment of metabolic and volumetric changes for assessment of tumor response in patients with soft-tissue sarcomas.* J Nucl Med, 2008. **49**(10): p. 1579-84.
27. Willett, C.G., et al., *Three-dimensional volumetric assessment of response to treatment: stage I and II diffuse large cell lymphoma of the mediastinum.* Radiother Oncol, 1988. **12**(3): p. 193-8.
28. Willett, C.G., et al., *Stage IA to IIB mediastinal Hodgkin's disease: three-dimensional volumetric assessment of response to treatment.* J Clin Oncol, 1988. **6**(5): p. 819-24.
29. Wakelee, H.A., et al., *Changes in the natural history of nonsmall cell lung cancer (NSCLC)--comparison of outcomes and characteristics in patients with advanced NSCLC entered in Eastern Cooperative Oncology Group trials before and after 1990.* Cancer, 2006. **106**(10): p. 2208-17.
30. Gandara, D.R., et al., *Radiographic imaging of bronchioloalveolar carcinoma: screening, patterns of presentation and response assessment.* J Thorac Oncol, 2006. **1**(9 Suppl): p. S20-6.
31. Goldstraw, P., et al., *The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours.* J Thorac Oncol, 2007. **2**(8): p. 706-14.
32. Jemal, A., et al., *Cancer statistics, 2008.* CA Cancer J Clin, 2008. **58**(2): p. 71-96.
33. Jaffe, C.C., *Measures of response: RECIST, WHO, and new alternatives.* J Clin Oncol, 2006. **24**(20): p. 3245-51.
34. Zhao, B., et al., *A Pilot Study of Volume Measurement as a Method of Tumor Response Evaluation to Aid Biomarker Development.* Clin Cancer Res, 2010. **16**: p. 4647-4653.
35. Fraioli, F., et al., *Volumetric evaluation of therapy response in patients with lung metastases. Preliminary results with a computer system (CAD) and comparison with unidimensional measurements.* Radiol Med, 2006. **111**(3): p. 365-75.
36. Parkin, D.M., et al., *Global cancer statistics, 2002.* CA Cancer J Clin, 2005. **55**(2): p. 74-108.
37. Llovet, J.M. and J. Bruix, *Systematic review of randomized trials for unresectable hepatocellular carcinoma: Chemoembolization improves survival.* Hepatology, 2003. **37**(2): p. 429-42.
38. Lopez, P.M., A. Villanueva, and J.M. Llovet, *Systematic review: evidence-based management of hepatocellular carcinoma--an updated analysis of randomized controlled trials.* Aliment Pharmacol Ther, 2006. **23**(11): p. 1535-47.
39. Koshariya, M., et al., *An update and our experience with metastatic liver disease.* Hepatogastroenterology, 2007. **54**(80): p. 2232-9.

40. Keil, S., et al., *Semi-automated measurement of hyperdense, hypodense and heterogeneous hepatic metastasis on standard MDCT slices. Comparison of semi-automated and manual measurement of RECIST and WHO criteria.* Eur Radiol, 2008. **18**(11): p. 2456-65.
41. Therasse, P., et al., *New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada.* J Natl Cancer Inst, 2000. **92**(3): p. 205-16.
42. Llovet, J.M., et al., *Sorafenib in advanced hepatocellular carcinoma.* N Engl J Med, 2008. **359**(4): p. 378-90.
43. Cheng, A.L., et al., *Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial.* Lancet Oncol, 2009. **10**(1): p. 25-34.
44. Garant, M., et al., *Liver metastasis: comparison of 2 methods for reporting of disease in patients receiving chemotherapy.* Can Assoc Radiol J, 1999. **50**(1): p. 13-6.
45. Stillwagon, G.B., et al., *194 hepatocellular cancers treated by radiation and chemotherapy combinations: toxicity and response: a Radiation Therapy Oncology Group Study.* Int J Radiat Oncol Biol Phys, 1989. **17**(6): p. 1223-9.
46. Sohaib, S.A., et al., *CT assessment of tumour response to treatment: comparison of linear, cross-sectional and volumetric measures of tumour size.* Br J Radiol, 2000. **73**(875): p. 1178-84.
47. Luccichenti, G., et al., *Radiologic assessment of rectosigmoid cancer before and after neoadjuvant radiation therapy: comparison between quantitation techniques.* AJR Am J Roentgenol, 2005. **184**(2): p. 526-30.
48. Rohde, S., et al., *CT-based evaluation of tumor volume after intra-arterial chemotherapy of locally advanced carcinoma of the oral cavity: comparison with clinical remission rates.* Cardiovasc Intervent Radiol, 2007. **30**(1): p. 85-91.
49. Therasse, P., *Measuring the clinical response. What does it mean?* Eur J Cancer, 2002. **38**(14): p. 1817-23.
50. McHugh, K. and S. Kao, *Response evaluation criteria in solid tumours (RECIST): problems and need for modifications in paediatric oncology?* Br J Radiol, 2003. **76**(907): p. 433-6.

Appendices

Appendix A: 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
- Conklin, J. ICON Medical Imaging
- Dorfman, GS. 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
- Goldmacher, G. ICON Medical Imaging
- 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, HJ. 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 University
- 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
- Silver, M. Toshiba Medical Research Institute – USA
- Steinmetz, N. Translational Sciences Corporation
- Sullivan, DC. RSNA Science Advisor and Duke University
- Tang, Y. CCS Associates
- Thorn, M. Siemens AG
- Vining, DJ. MD Anderson Cancer Center
- 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.

Appendix B: Background Information

B.1 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. Purchasers can specify conformance with appropriate QIBA Profiles as a requirement in Requests For Proposals (RFPs). Vendors who have successfully implemented QIBA Profiles in their products can publish QIBA Conformance Statements. The Conformance Statements are accompanied by “Model-specific Parameters” (as shown in Appendix D) describing how to configure their product for alignment with the Profile.

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.

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]).

B.2 CT Volumetry for Cancer Response Assessment: Overview and Summary

Anatomic imaging using computed tomography (CT) has been historically used to assess tumor burden and to determine tumor response to treatment (or progression) 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 tumors 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 [1, 2]. 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 tumors 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 tumor [3]. RECIST defines response as a 30% decrease of the largest diameter of the tumor. For a spherical tumor, 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 linear tumor diameters as defined by RECIST. Comparative analyses in the context of clinical trial data have found volume measurements to be more reliable, and often more sensitive to longitudinal changes in size and thus to treatment response, than the use of a uni-dimensional diameter 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 [4-7].

The rationale for volumetric approaches to 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 believed to be 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% decrease in tumor volume. If one assumes a 21 mm diameter spherical tumor (of 4.8 cc volume), partial response would require that the tumor shrink to a diameter of less than 15 mm, which would correspond to a decrease in volume all the way down to 1.7 cc. The much greater absolute magnitude of volumetric changes is potentially less prone to measurement error than changes in diameter, particularly if the tumors are spiculated or otherwise irregularly shaped. 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.

848 **Table B.1 Summarizing the precision/reproducibility of volumetric measurements from clinical studies reported in the literature**

Scan	Reader	# of Readers	# of Patients	# of Nodules	Tumor 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 [8]
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 [9]
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 [10]
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 [10]
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 [10]
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 [10]
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 [11]
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 [13]
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 [14]
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 [14]
same scan	inter-reader	2	2239	4225	15–500 mm3 (effective diameter 3.1–9.8 mm)	lung, nodules	-13.4 to 14.5%	0.50%	not reported	not reported	1.0/0.7	Wang *et al*. 2008 [15]

Scan	Reader	# of Readers	# of Patients	# of Nodules	Tumor 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
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*. [16]
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*. [16]

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

The above table provides a basis for the 30% value in the Profile Claim. The range between the minimum and maximum values in the 95% CI of the measurement difference column is mostly within +/- 15%. Considering a large study from Wang et al using 2239 patients [15], the 95% confidence interval ranged [-13.4%, 14.5%]. Thus, 30% is a conservative threshold of measurement variation. For example, the 30% change in the claim is the outside of 95% confidence interval of 15% of measurement variability when sample size is 40 or more.

B.3 Detailed Literature Review by Indication

To date, volumetry has been evaluated in lung, liver, head and neck, esophagus, and rectal cancers, sarcoma, and lymphoma (Appendix 1, Tables 1–7). Most studies compared volumetry with either unidimensional RECIST or bidimensional WHO classifications. Volumetry showed a high degree of concordance with uni- or bidimensional assessment in some studies [17, 18]; others showed considerable discordance between these methods in response classifications [19-22]. Correlation of volumetric assessment with pathologic response was examined in four studies (two esophageal, one gastric cancer, and one sarcoma) in patients who had or were having neoadjuvant chemotherapy. Among those four studies, volumetric assessment was correlated with pathologic response in two studies (one esophageal and one gastric study) [23, 24], whereas no such correlation was found in an esophageal study [25] and a sarcoma study [26]. Two of the above neoadjuvant studies also followed esophageal cancer patients for OS or PFS, but neither study showed correlation with volumetric assessment [24, 25]. In addition, two small studies [27, 28] with lymphoma patients showed that patients with greater reduction in tumor volume at 1–2 months of chemotherapy had lower probability of relapse at one year.

Lung Cancer (Tables B.2 and B.3)

Lung cancer typically spreads as it advances from localized disease to the neighboring lymph nodal structures of the lung (regional metastatic spread). The most advanced stage is metastasis to a distant site such as the brain or liver. In clinical trials, depending on the initial stage at diagnosis, either progression of localized disease or the discovery of a new site of metastatic dissemination is the basis for declaring failure of the efficacy of a new drug. In virtually all lung cancer clinical trials, there are situations when either a quantitative or a qualitative endpoint may be relevant, but it is likely that quantitative endpoints will be most frequently informative in trials.

With advanced disease, there is a tendency toward more frequent disease progression at a distant metastatic site rather than progression due to extension from the primary tumor [29]. These patterns of disease progression impact clinical trial design in measuring drug response. However, there are exceptions to the pattern just described, such as bronchioalveolar carcinoma. This more indolent cancer tends to spread extensively within the lung but seldom to distant sites [30].

Staging, Therapeutic Options, and Response Assessment by Imaging Approaches

Staging defines the extent of lung cancer dissemination at the time of initial patient diagnosis. The schema for staging lung cancer has been updated recently to more accurately cluster patients who benefit from particular therapeutic interventions with predictable outcomes [31]. How staging relates to lung cancer drug therapy approaches, the imaging approaches used in those stages, and issues relative to image requirements is summarized in Table B.2 [32].

Table B.2. Summary of Image Processing Issues Relative to Lung Cancer Stage

Stage	Percent of Cases	Percent 5-year Survival	Imaging Focus/ Therapy Focus	Imaging Tool	Issues	Thoracic Segmentation	Hi-Res
I	16	49	Primary tumor/ Neo and adjuvant RX	MDCT	Small cancers surrounded by air	Can be straightforward	Needed
II/III	35	15.2	Primary, hilar, and mediastinal lymph nodes/Combined modality	MDCT, PET	Larger tumors and nodes abut other structures	Often challenging	Optional
IV	41	3	Primary/regional nodes and metastatic sites/ Chemotherapy	MDCT, PET, bone, brain scans	Tumor response often determined outside of the chest	Often challenging	Optional

The imaging goal may vary in different disease stages. For example, with Stage IV lung cancer, the disease progression could be due to new growth in the primary lung tumor and/or metastasis of the cancer to a distant site, and not growth of the primary cancer site. In Stage II and III lung cancer, disease progression is often manifested by increased tumor involvement in regional lymph nodes. CT imaging would typically be used to assess potential disease progression in either the primary tumor or in the lymphatic tissue. The development of new sites of metastatic disease in a Stage IV clinical trial will require a different imaging approach. To assess for new sites of metastatic disease, CT may be used to look for thoracic, hepatic, or retroperitoneal sites of metastasis, and PET scans will frequently be used to assess the progression of metastatic disease across the entire body. Common both to improving size-based measures (*i.e.*, moving from linear diameters to volume) as well as more computationally sophisticated measures (*e.g.*, tissue density in CT, mechanistic measures in PET) is a need for means to qualify performance across stakeholders involved in the application of these measures.

The potential utility of volumetry in predicting treatment response in lung cancer patients has been investigated by several groups. Jaffe pointed out that the value of elegant image analysis has not been demonstrated yet in clinical trials [33]. Value depends, at least in part, on the extent to which imaging endpoints meet criteria as substitute endpoints for clinical outcome measures. In this review, however, value is limited to the ability of imaging to predict either beneficial biological activity or progressive disease sooner than alternative methods of assessment, so that individual patients can move on to other treatment alternatives, or at the very least, stop being exposed to toxicity without benefit. In this context, value is predominantly a function of sensitivity and accuracy.

In 2006, Zhao and colleagues [4] reported a study of 15 patients with lung cancer at a single center. They used MDCT scans with a slice thickness of 1.25 mm to automatically quantify unidimensional LDs, bidimensional cross products, and volumes before and after chemotherapy. They found that 11/15 (73%) of the patients had changes in volume of 20% or more, while only one (7%) and 4 (27%) of the sample had changes in uni- or bidimensional line-lengths of >20%. Seven (47%) patients had changes in volume of 30%

or more; no patients had unidimensional line-length changes of 30% or more, and only two patients (13%) had changes in bidimensional cross products of 30% or more. The investigators concluded that volumetry was substantially more sensitive to drug responses than uni- or bidimensional line-lengths. However, this initial data set did not address the clinical value of increasing the sensitivity of change measurements.

In a follow-up analysis [34], the same group used volumetric analysis to predict the biologic activity of epidermal growth factor receptor (EGFR) modulation in NSCLC, with EGFR mutation status as a reference. In this population of 48 patients, changes in tumor volume at three weeks after the start of treatment were found to be more sensitive and equally specific when compared to early diameter change at predicting EGFR mutation status. The positive predictive value of early volume response for EGFR mutation status in their patient population was 86%. The investigators concluded that early volume change has promise as an investigational method for detecting the biologic activity of systemic therapies in NSCLC.

In 2007, Schwartz and colleagues [6] unidimensionally and volumetrically evaluated target lesions, including lymph node, liver, peritoneal, and lung metastases, in 25 patients with metastatic gastric cancer being treated with combination therapy, and reported that volumetry predicted clinical response earlier than unidimensional RECIST by an average of 50.3 days.

In 2008, Altorki and colleagues [7] reported that volumetry is substantially more sensitive than changes in unidimensional diameters. In a sample of 35 patients with early-stage lung cancer treated with pazopanib, 30 of 35 (85.7%) were found to have a measurable decrease in tumor volume; only three of these 35 subjects met RECIST criteria for a PR.

In a retrospective analysis of 22 patients with locally advanced lung cancer treated with radiation and chemotherapy, assessment of treatment response by volume change was found to be in agreement with that by RECIST and WHO criteria (K 0.776; 95% CI 0.357–1.0 for agreement with both RECIST and WHO) [18] in 21 of 22 patients.

In another retrospective analysis of 15 patients with lung metastases from colorectal cancer, renal cell, or breast carcinoma, volumetric assessment of 32 lung lesions at baseline and after 1–4 months standard chemotherapy or radiotherapy showed fair to poor agreement with either RECIST or WHO assessment for response classification [19].

In another retrospective analysis of 68 patients with primary or metastatic lung malignancies, volumetric assessment of treatment response was found to be highly concordant with RECIST (K 0.79–0.87) and WHO assessment (K 0.83–0.84) [17]. The intraobserver reproducibility of volumetric classification was 96%, slightly higher than that of RECIST and WHO. The relative measurement error of volumetric assessment was 8.97%, also slightly higher than that of unidimensional and bidimensional assessment.

In another retrospective analysis of nine patients with lung metastases who were undergoing chemotherapy, volumetric assessment of treatment response agreed in all but one case with RECIST assessment at the patient level (K 0.69); at the lesion level, volumetric and RECIST assessment agreed on 21 of the 24 lesions (K 0.75). The level of agreement between volumetric and RECIST assessment was equivalent or superior to that of inter-observer agreement using the RECIST criteria [35].

Primary Liver Cancer and Metastatic Lesions in the Liver (Table B.4)

Hepatocellular carcinoma (HCC) is the most common form of liver cancer in adults [36]. The majority of patients have underlying hepatic dysfunction, which complicates patient management and trial design in the search for effective treatment [37, 38]. Despite advances in many aspects of HCC treatment, >70% of HCC patients present with advanced disease and will not benefit from existing treatment modalities, including liver transplantation, surgical resection, and loco-regional therapies. At present, only one systemic agent, *i.e.*, sorafenib, is approved for advanced HCC patients. There remains a great need for safe and effective systemic therapies for HCC patients who progressed on or do not tolerate sorafenib and for patients with more advanced hepatic dysfunction. The liver is also a common site of metastatic spread; metastatic involvement of the liver can occur with many neoplasms, including lung, colorectal, esophageal, renal cell and breast, and stomach cancers, pancreatic carcinoma, and melanoma [39, 40].

Evidence that radiologic responses reflect clinical outcomes has recently emerged in patients who were receiving systemic therapy for advanced liver cancer. In a phase 3 trial, sorafenib, a small molecule kinase inhibitor, prolonged the survival of patients with advanced liver cancer to 10.7 months as compared with 7.9 months for the placebo group. The time to radiologic progression as defined by RECIST [41] was also significantly prolonged in the sorafenib group, in parallel with the survival advantage [42]. This survival advantage conferred by sorafenib was later confirmed in the Asian population [43].

Volumetric CT has been investigated in only a few studies in patients with metastatic liver lesions [21, 44] or HCC [45] (Appendix 1) as discussed below. These studies compared volumetry with RECIST and/or the bidimensional WHO method in classifying treatment response, and found considerable discordance between volumetry and RECIST or WHO assessment [21, 44].

Prasad and colleagues [21] compared volumetric with unidimensional (RECIST) and bidimensional (WHO) measurements in assessing response to treatment in 38 patients with liver metastases from breast cancer in a phase 3 trial. PR was defined as >65% reduction in volume; PD was defined as >73% increase in volume; and stable disease was defined as changes in volume between those in PR and PD. Patients were treated with docetaxel or capecitabine plus docetaxel, and tumors were measured at baseline and six months posttreatment. Response assessment using uni- and bidimensional methods are highly concordant (37 of 38 patients). Volumetric assessment of tumor burden was discordant with uni- and bidimensional results in 12 (32%) and 13 (34%) patients, respectively.

In another retrospective analysis of 10 patients with liver metastases from colorectal (8), esophageal (1), and gastric (1) cancers who were receiving chemotherapy, 26 pairs of pre- and posttreatment CT scans were evaluated by bidimensional criteria (WHO) and volumetry. Stable disease in the volumetric analysis was defined as between an increase in volume of less than 40% and a reduction in volume of less than 65%. Discordance between the bidimensional assessment and volumetry was found in 19–35% of the cases in disease status categories [44].

Stillwagon and colleagues [45] used volumetric measurements to assess the response to radiation and chemotherapy in 194 patients with unresectable HCC. PD was defined as 25% increase in volume; PR was defined as 30% reduction in volume; and stable disease was defined as less than 25% increase or less than 30% decrease in tumor volume.

Lymphoma (Table B.5)

Lymphomas comprise ~30 distinct diseases. Volumetric assessment of lymphoma has been found to correlate with treatment outcome in two early studies [27, 28] using non-helical scanners. Agreement with RECIST and WHO assessment was also found to be excellent in another study [46].

In a study of eight patients with Stage I and II diffuse large cell lymphoma of the mediastinum followed for 12 to 68 months (mean 29 months), tumor volume was assessed before and at 1 to 2 months after chemotherapy. The relative tumor volume reduction was higher in those who remained in remission than in patients who had relapsed (89% and 73% reduction, respectively). However, whether this difference was statistically significant was not reported. It was also noted that the initial tumor volume prior to chemotherapy was also greater in the group who later relapsed [27].

In a study of 12 patients with stage IA to IIB mediastinal Hodgkin's disease who were followed for 12 to 84 months (mean 35 months) after treatment, patients with a >85% reduction in volume at 1 to 2 months after six cycles of chemotherapy had a lower incidence of mediastinal relapse (0/6, 0%) compared with those having 85% of less reduction (4/6, 67%) [28].

In a study of 16 patients with lymphoma or germ cell tumors, volumetric assessment of response to chemotherapy agreed completely with the WHO criteria in classifying responses of the lesions (20 lesions), and agreed in 18 of the 20 (90%) lesions with RECIST criteria [46].

Colorectal and Gastric Cancers (Table B.6)

Data suggest that volumetry may be valuable in assessing response to neoadjuvant therapy in gastric and colorectal cancers. In a prospective phase 2 study in 33 patients with resectable advanced gastric cancer who had four cycles (eight weeks) of neoadjuvant chemotherapy before surgical resection, volume reduction of primary gastric cancer correlated with histopathologic grades of regression, but the unidimensional reduction of maximum thickness and standardized uptake value (SUV) of FDG-PET did not. The optimal cut-off value of the tumor volume reduction was determined to be 35.6%, resulting in a positive predictive value and negative predictive value of 69.9% and 100%, respectively [23].

In a study of 15 patients with rectosigmoid cancer prospectively enrolled in neoadjuvant radiation therapy, using a reduction of >65% in tumor volume as the threshold for PR, volumetric analysis disagreed with the WHO criteria in classifying treatment response in one patient and with the RECIST assessment (measuring the maximal wall thickness) in four patients [47].

Head and Neck Cancer (Table B.7)

Head and neck cancers are clinically heterogenous, comprising multiple anatomic sites of origin with distinct natural histories and prognoses. Cure rates are low (30–50%) in locally advanced disease.

The role of volumetry in response assessment in head and neck cancer is unclear. In two retrospective studies of 129 patients with early or late stages of oral cavity or oropharynx carcinoma, assessment of

response by volumetry had low agreement (38–56%) with clinical assessment by inspection and palpation [22, 48]. In the first study of 42 patients with early-stage oral cavity or oropharynx carcinoma, volume assessment of response at three to four weeks after local chemotherapy had low agreement with clinical assessment by inspection and palpation according to WHO criteria (38%) in classifying treatment response. It is noted that the lesion volume was calculated manually, assuming lesions were ellipsoid-shaped [22].

In the second retrospective study reported by the same group, 87 patients with advanced oral cavity or oropharynx carcinoma were assessed by lesion volume before and three weeks after local chemotherapy. Volume assessment of treatment response agreed with clinical assessment by WHO criteria in 49 of 87 patients (56%) [48].

Sarcoma (Table B.8)

The response to treatment in sarcoma is difficult to objectively measure and quantify anatomically as shown by the limited usefulness of RECIST in this setting [49, 50]. Assessment of tumor dimensions in sites such as bone, bowel, and peritoneal metastases is problematic; in addition, tumor volume reductions that can be measured by standard criteria may occur slowly or not at all (*e.g.*, due to persistence of necrotic or fibrotic tissue).

Volumetry has not demonstrated a value in response assessment in sarcoma. In a study of 20 patients with locally advanced high-grade soft-tissue sarcoma prospectively enrolled in neoadjuvant therapy, volume assessment before and after pre-operative treatment failed to correlate with histopathologic response and was unable to differentiate histopathologic responders (n=6) from non-responders (n=14). In contrast, changes in FDG uptake measured by SUV (both mean and maximum) using PET were predictive of histopathologic response at a high accuracy (area under response operating characteristics (ROC) curve = 1.0 and 0.98, respectively) [26].

1089
1090

Table B.3. Evaluation of Response to Therapy by Volumetry in Lung Cancer

Disease Stage/ Therapy	Number of Patients Evaluated	VIA Response Measurement/Timing	Comparator	Results	Statistical Analysis	Reference
NSCLC, locally advanced/ radio ± chemo (mostly carboplatin/ paclitaxel	22	PR –65%	RECIST, WHO	Good concordance between 3D, 2D, 1D (cases). CR 4/4/4, PR 16/15/15, NR 2/3/3.	Kappa values. 3D *vs* 2D Kappa 0.776 (95% CI 0.357–1.0, substantial agreement); 3D *vs* 1D Kappa 0.776 (95% CI 0.357–1.0, substantial agreement); 1D *vs* 2D Kappa 1.0 (perfect agreement)	Werner-Wasik *et al*. 2001[18]
NSCLC, early stage gefitinib 3 wks, neoadjuvant	48	–24.9% (dichotomizing cut-off)	EGFR mutation sensitizing tumor to tyrosine kinase inhibitor; volume change -65% (RECIST deduced); optimal cut-off 1D (–7%)	Optimal cut-off of 3D changes 24.9%; sensitivity 90%, specificity 89% for classifying tumor w/o EGFR sensitizing mutation; PPV 86%, NPV 92%. 3D (24.9%) superior to 1D (optimal and RECIST).	Youdens' index (sensitivity + specificity −1) for determination of optimal dichromatic cut-off value; Wilcoxon rank-sum test for significance of difference	Zhao *et al* 2010 [34]
Lung mets from colorectum, renal cell, breast; standard chemo or radio	15	Stable disease -65% to +44%; 2 follow-ups, at 1–4 months	RECIST, WHO	Kappa 3D *vs* 1D 0.818 (Visit 1 to V2), 0.429 (V2 to V3); 3D *vs* 2D 0.412 (V1 to V2), 0.118 (V2 to V3); fair agreement 3D *vs* 1D; poor 2D *vs* 3D	Kappa values	Tran *et al* 2004 [19]
NSCLC (16), SCLC (9), lung mets of various origins (43); treatment not specified	68	Stable disease –65% to +44%; 3 months for lung cancer, time varied for mets	RECIST, WHO	Kappa 1D *vs* 3D 0.79-0.87, Kappa 2D *vs* 3D 0.83-0.84	Kappa values	Sohns *et al*. 2010 [17]
Lung mets, unspecified origin; chemo	9 (24 nodules)	Stable disease –65% to +73%;	RECIST	At nodule/lesion level, disagreement 3 in 24 nodules (Kappa 0.75); at patient level, disagree 1/9 (Kappa 0.59)	Kappa values	Fraioli *et al*. 2006 [35]

NSCLC, stage I or II, operable and resectable/ gefitinib > 21 days	15	–20% and –30%; 26.4 days since baseline scan	RECIST and WHO	3D more sensitive in detecting changes. > –20%: 3D: 11/15 (73%); 1D 1/15 (7%) (p< .01); 2D 4/15 (27%)(P= .04); > –30%: 3D, 7/15 (47%); 1D 0/15 (p= .02); 2D, 2/15 (13%) (p= .06).	P values	Zhao *et al*. 2006 [4]
Mets to lymph node, liver, peritoneal and lung originated from primary gastric cancer or Gastroespoha-geal junction adenocarcinoma/ irinotecan, cisplatin and bevacizumab	25	3D, –65%/ 6-week follow-up for 10 cycles. 1D and 3D comparison made at the time with maximal clinical response	RECIST	8/25 (72%) responders by both RECIST and 3D; 3D identified responders a mean of 50.3 days earlier than did RECIST criteria	There was a statistically significant (p<0.01) change in ratio of volume measurement change to RECIST measurement change for responding versus stable patients.	Schwartz *et al.* 2007 [6]
NSCLC, Stage I/II Resectable/ neoadjuvant, pazopanib 800mg qd for 2 to 6 weeks	35	Volume change, response criteria not specified/1 week after last dose	RECIST	3D: 30/36 (86%) –1% to –86%. 2/35 > –50% (86% and 75%; 23/35 (66%) > –10%; 12/35 > –30%; 1D 3/25 PR (reduction 86%, 75%, and 36%). Discordance between 3D and RECIST, not head-to-head comparison in % change. 3D superiority unclear.	Not specified	Altorki *et al.* 2010 [7]

1093 **Table B.4. Evaluation of Response to Therapy by Volumetry in Liver Cancer**
1094

Disease Stage/ Therapy	Number of Patients Evaluated	VIA Response Measurement/Timing	Comparator	Results	Statistical Analysis	Reference
Hepatic mets from GI systemic chemo +/- intra-arterial chemo	10 (37 scans)	Stable disease –65% to +40%	WHO	Discordance between 2D and 3D: 19% and 35% (w/o or w 5% variation interval). Conclusion: 2D and 3D not interchangable. 2D tended to identify PD when 3D indicated no change. All lesions hypodense.	Not specified	Garant *et al.* 1999 [44]
Hepatic mets from breast docetaxel *vs* capecitabine + docetaxel	38	Stable disease –65% to +73%	RECIST, WHO	Treatment response concordance 1D and 2D; discordance 1D *vs* 3D, and 2D *vs* 3D	Not specified	Prasad *et al.* 2000 [21]

1095
1096 **Table B.5. Evaluation of Response to Therapy by Volumetry in Lymphoma**
1097

Disease Stage/ Therapy	Number of Patients Evaluated	VIA Response Measurement/Timing	Comparator	Results	Statistical Analysis	Reference
Lymphoma or germ cell; chemo	16 (20 lesions)	Volume change/timing not specified	RECIST, WHO	3D agreed completely with 2D, agreed in 18/20 (lesions) with 1D	Coefficient of variation calculated	Sohaib *et al.* 2000 [46]
Diffuse large cell lymphoma of the mediastinum; multiagent chemo	8	Volume change; 1–2 months (CT follow-up)	Relapse/ remission/ death	Patients were followed for minimum 1 yr or until death, mean 29 months (13–68 months). Reduction of tumor volume greater in pts in remission than in relapse (89% *vs* 73%, respectively).	No statistical analysis performed	Willett *et al.* 1988 [27]
Mediastinal Hodgkin's, stage IA to IIB; multiagent chemo	12	Volume change; 1–2 months (CT follow-up)	Relapse/ remission/ death	Patients were followed for minimum 1 yr or until death, mean 35 months (12–84 months). a >85% reduction in volume at 1 to 2 months after six cycles of chemotherapy had a lower incidence of mediastinal relapse (0/6, 0%) compared with those having 85% of less reduction (4/6, 67%)	No statistical analysis performed	Willett *et al.* 1988 [28]

1098
1099

1100 **Table B.6. Evaluation of Response to Therapy by Volumetry in Colorectal and Gastric Cancers**
1101

Disease Stage/ Therapy	Number of Patients Evaluated	VIA Response Measurement/Timing	Comparator	Results	Statistical Analysis	Reference
Gastric, resectable advanced; chemo, neoadjuvant: oxaliplatin, infusional 5-FU, leucovorin; 8 wks (4 cycles)	33	Volume change cut-off 35.6%; 4, 8 wks	Pathologic response	Volume reduction of primary tumor or index node correlated best with histopathologic grades for regression, followed by short diameter of index node. No correlation with thickness of primary gastric cancer, or SUV by PET/CT. The optimal cut-off value of the tumor volume reduction 35.6%.	Spearman rank for correlation determination; ROC for determination of optimal cut-off value	Lee *et al.* 2009 [23]
Rectosigmoid; neoadjuvant radiation	15	PR –65%; timing not specified	Maximal wall thickness (RECIST), WHO	Discordance w RECIST and WHO (4/15 and 1/15, respectively)	Student's t test for paired data; Pearson's correlation test. $p < 0.05$	Luccichenti *et al.* 2005 [47]

1102
1103
1104 **Table B.7. Evaluation of Response to Therapy by Volumetry in Head and Neck Cancer**
1105

Disease Stage/ Therapy	Number of Patients Evaluated	VIA Response Measurement/Timing	Comparator	Results	Statistical Analysis	Reference
Oral cavity and oropharynx, carcinoma T1/2; chemo (cisplatin), intra-arterial	42	CR –90%, PR -50%, stable disease –50% to +25%, PR >+25%; 4 wks	Clinical inspection and palpation of lesions, classified per WHO criteria	Concordance in classifying response categories 16 of 42 pts (38%)	Not reported for concordance analysis	Rohde 2006 [22]
Oral cavity and oropharynx, carcinoma T3/4; chemo (cisplatin), intra-arterial	87	CR –90%, PR –50%, stable disease –50% to +25%, PR >+25%; 4 wks	Clinical inspection and palpation of lesions, classified per WHO criteria	Concordance in classifying response categories 49/87 pts (56%); Kappa value was not reported.	Kappa for agreement between clinical and radiological remission rates	Rohde 2007 [48]

1106
1107

Table B.8. Evaluation of Response to Therapy by Volumetry in Sarcoma

Disease Stage/ Therapy	Number of Patients Evaluated	VIA Response Measurement/Timing	Comparator	Results	Statistical Analysis	Reference
Sarcoma, locally advanced high-grade, soft tissue; chemo (ifosfamide/doxorubicin or gemcitabine/docetaxel)± radiation	20	Volume change/timing not specified	Pathologic response	Volume reduction not significant pre- and post-treatment; not predictive of histopathologic response (6 responders, 14 nonresponders), AUC = 0.48	ROC curve	Benz 2008 [26]

Abbreviations:
1D = unidimensional measurement; 2D = bidimensional measurement; 3D = volumetric measurement; AUC = area under the curve; CI = confidence interval; CR = complete response; EGFR = epidermal growth factor receptor; FU = fluorouracil; Mets = metastasis; NSCLC = non small cell lung cancer; OS = overall survival; PFS = progression free survival; PR = partial response; PR = partial response; RECIST = Response Evaluation Criteria in Solid Tumors; ROC = response operating characteristics; SCLC = small cell lung cancer.

Appendix C: 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.)

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: that part of the image 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.

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

Tumor Definition Variability: the clarity of the tumor boundary in the images. It originates from the biological characteristics of the tumor, technical characteristics of the imaging process, and perhaps on the perception, expertise and education of the operator.

Technical Variability - originates only from the ability to drawing unequivocal objects. In other words, the perception of tumor definition is supposed absolutely clear and similar for any given operator when attempting to assess “Technical” variability.

Global Variability - partitioned as the variability in the tumor definition plus the “Technical” variability.

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.

Repeatability – considers multiple measurements taken under the same conditions (same equipment, parameters, reader, algorithm, etc) but different subjects.

Reproducability – considers multiple measurements taken where one or more conditions have changed.

Appendix D: 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 3.2, 3.3 and 3.4.

This Appendix provides, as an informative annex to the Profile, 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.

Additional parameter sets may be found in QIBA Conformance Statements published by vendors and sites. Vendors claiming product compliance with this QIBA Profile are required to provide such instructions and parameters describing the conditions under which their product achieved compliance.

Sites using models listed here are encouraged to consider 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. Tables like the following may be used by sites that wish to publish their successful/best practices.

In any case, sites are responsible for adjusting the parameters as appropriate for individual subjects.

Discussion:
It would likely be useful to include a description of the imaging subject in the following tables.
In terms of standardization, it may make sense to ask vendors to publish parameters for a known reference phantom as a stable benchmark for sites to adjust for individual patient variations.

Table D.1 Model-specific Parameters for Acquisition Devices

Acquisition Device	Settings Compatible with Compliance	
<Vendor> <Model> <Version>	Submitted by:	
	kVp	
	Number of Data Channels (N)	
	Width of Each Data Channel (T, in mm)	
	Gantry Rotation Time in seconds	
	mA	
	Pitch	
	Scan FoV	

1175 **Table D.2 Model-specific Parameters for Reconstruction Software**

Reconstruction Software	Settings Compatible with Compliance	
<Vendor> <Model> <Version>	Submitted by:	
	Reconstructed Slice Width, mm	
	Reconstruction Interval	
	Display FOV, mm	
	Recon kernel	

1176

1177 **Table D.3 Model-specific Parameters for Image Analysis Software**

Image Analysis Software	Settings Compatible with Compliance	
<Vendor> <Model> <Version>	Submitted by:	
	a	
	b	
	c	
	d	

1178