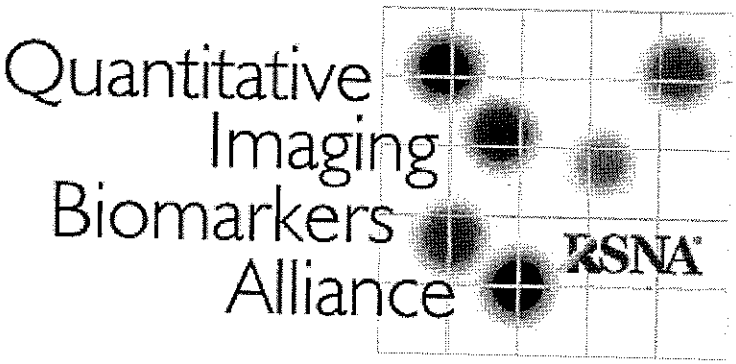

QIBA Profile:

CT ~~Tumor~~ Volume Change (CTV-1)

IN EVALUATING PRIMARY Lung Cancer

Version 2.2
8 Aug 2012
Status: (pre)Reviewed

Table of Contents

1. Executive Summary 6
2. Clinical Context and Claims 7
3. Profile Details 7
3.1. Subject Handling 9
3.2. Image Data Acquisition 12
3.3. Image Data Reconstruction 15
3.4. Image Analysis 17
4. Compliance 21
4.1. Performance Assessment: Tumor Volume Change Variability 21
4.2. Performance Assessment: Image Acquisition Site 23
References 26
Appendices 29
Appendix A: Acknowledgements and Attributions 29
Appendix B: Background Information 30
Appendix C: Conventions and Definitions 46
Appendix D: Model-specific Instructions and Parameters 47

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.

Clinician imaging screening process

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~~ 5 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.

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.

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

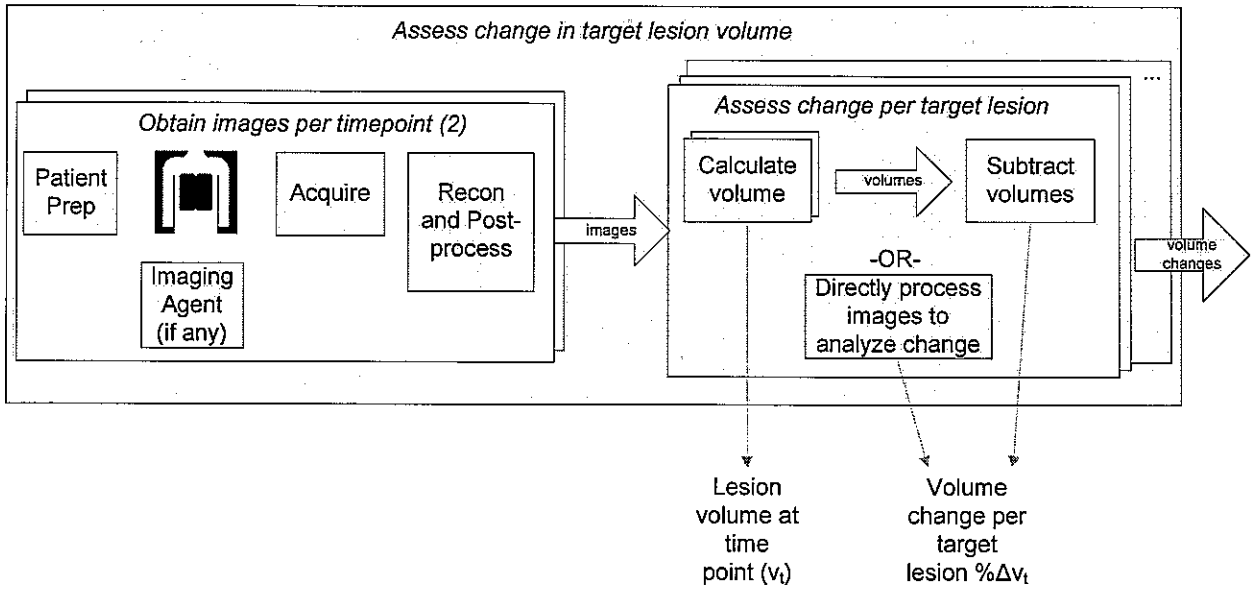

Figure 1: CT Tumor Volumetry - Activity Sequence

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

3
4

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, retroperitoneal, 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."

4.2.1 Acquisition Validation

Review patient handling procedures for compliance with Section 3.1

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

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

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

4.2.2 Test Image Set

Locally acquire a test image set using the protocols established and tested in Section 4.2.1.

The test image set should conform to the characteristics described in Section 4.1.1.

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