

QIBA Profile:

CT Tumor Volume Change for Advanced Disease (CTV-AD)

Stage: Publically Reviewed (Draft)

When referencing this document, please use the following format:

CT Volumetry Technical Committee. CT Tumor Volume Change Profile - 2016, Quantitative Imaging Biomarkers Alliance. Publically Reviewed Draft. QIBA, May 1, 2016. Available at: http://rsna.org/QIBA\_.aspx

Table of Contents

[1. Executive Summary 4](#_Toc449269951)

[2. Clinical Context and Claim(s) 5](#_Toc449269952)

[3. Profile Requirements 9](#_Toc449269953)

[3.1. Periodic QA 11](#_Toc449269954)

[3.1.1 Discussion 12](#_Toc449269955)

[3.1.2 Specification 12](#_Toc449269956)

[3.2. Subject Handling 12](#_Toc449269957)

[3.2.1 Discussion 12](#_Toc449269958)

[3.2.2 Specification 14](#_Toc449269959)

[3.3. Image Data Acquisition 16](#_Toc449269960)

[3.3.1 Discussion 16](#_Toc449269961)

[3.3.2 Specification 18](#_Toc449269962)

[3.4. Image Data Reconstruction 20](#_Toc449269963)

[3.4.1 Discussion 20](#_Toc449269964)

[3.4.2 Specification 22](#_Toc449269965)

[3.5. Image QA 24](#_Toc449269966)

[3.5.1 Discussion 24](#_Toc449269967)

[3.5.2 Specification 26](#_Toc449269968)

[3.6. Image Analysis 27](#_Toc449269969)

[3.6.1 Discussion 27](#_Toc449269970)

[3.6.2 Specification 29](#_Toc449269971)

[4. Assessment Procedures 32](#_Toc449269972)

[4.1. Assessment Procedure: In-plane Spatial Resolution 32](#_Toc449269973)

[4.2. Assessment Procedure: Voxel Noise 34](#_Toc449269974)

[4.3. Assessment Procedure: Tumor Volume Computation 34](#_Toc449269975)

[4.4. Assessment Procedure: Tumor Volume Change Repeatability 35](#_Toc449269976)

[4.4.1 obtain test image set 35](#_Toc449269977)

[4.4.2 determine volume change 36](#_Toc449269978)

[4.4.3 calculate statistical metrics of performance 36](#_Toc449269979)

[4.5. Assessment Procedure: Tumor Volume Bias and Linearity 37](#_Toc449269980)

[4.5.1 obtain test image set 37](#_Toc449269981)

[4.5.2 determine volume change 39](#_Toc449269982)

[4.5.3 calculate statistical metrics of performance 39](#_Toc449269983)

[4.X. Assessment Procedure: Image Acquisition Site Performance 40](#_Toc449269984)

[4.X.1 Acquisition Validation 41](#_Toc449269985)

[4.X.2 Test Image Set 41](#_Toc449269986)

[Closed Issues: 43](#_Toc449269987)

[Appendices 48](#_Toc449269988)

[Appendix A: Acknowledgements and Attributions 48](#_Toc449269989)

[Appendix B: Background Information 49](#_Toc449269990)

[Appendix C: Conventions and Definitions 50](#_Toc449269991)

[Appendix D: Model-specific Instructions and Parameters 51](#_Toc449269992)

[Appendix E: Metrology Definitions and Methods 53](#_Toc449269993)

# 1. Executive Summary

The goal of a QIBA Profile is to help achieve a useful level of performance for a given biomarker.

The **Claim** (Section 2) describes the biomarker performance.  
The **Activities** (Section 3) contribute to generating the biomarker. Requirements are placed on the   
**Actors** that participate in those activities as necessary to achieve the Claim.   
**Assessment Procedures** (Section 4) for evaluating specific requirements are defined as needed.

This QIBA Profile (CT Tumor Volume Change) addresses tumor volume change which is often used as a biomarker of disease progression or response to treatment. It places requirements on actors (Acquisition Devices, Technologists, Physicists, Radiologists, Reconstruction Software and Image Analysis Tools) involved in activities (Periodic QA, Subject Handling, Image Data Acquisition, Image Data Reconstruction, Image QA and Image Analysis).

The requirements are primarily focused on achieving sufficient accuracy and avoiding unnecessary variability of the tumor volume measurements.

The clinical performance target is to achieve a 95% confidence interval for the tumor volume change with precision of -25% to +30%.

This document is intended to help clinicians basing decisions on this biomarker, imaging staff generating this biomarker, vendor staff developing related products, purchasers of such products and investigators designing trials with imaging endpoints.

Note that this document are only states 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.

QIBA Profiles addressing other imaging biomarkers using CT, MRI, PET and Ultrasound can be found at qibawiki.rsna.org.

# 2. Clinical Context and Claim(s)

****Clinical Context****

Quantifying the volumes of tumors and measuring tumor longitudinal changes within subjects (i.e. evaluating growth or regression with image processing of CT scans acquired at different timepoints).

**Compliance with this Profile by all relevant staff and equipment supports the following claims:**

Claim 1: A true change in a tumor volume has occurred with 95% confidence if the measured change is larger than 24% and the longest in-plane diameter is initially 50-100mm.

Claim 2: A true change in a tumor volume has occurred with 95% confidence if the measured change is larger than 29% and the longest in-plane diameter is initially 35-49mm.

Claim 3: A true change in a tumor volume has occurred with 95% confidence if the measured change is larger than 39% and the longest in-plane diameter is initially 10-34mm).



**Claim 4:  The tumor volume measurement performance, expressed as within-subject coefficient of variation (wCV), is 0.157, 0.241, and 0.501 respectively for tumors with diameters of 50-100mm, 35-49mm, and 10-34mm. The resulting 95% confidence interval for the true change in volume for several example measured tumors is:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Baseline Diameter  ( Volume)** | **Subsequent Diameter (Volume)** | **Volume Change** | **95% Confidence** | **95% Confidence Interval of True Volume Change** |
| **100mm (523,600mm3)** | **50mm  (65,450mm3)** | **-458,150 mm3** | **±162,356 mm3** | **[-620,525 mm3, -295,773 mm3]** |
| **40mm  (33,510mm3)** | **80mm  (268,080mm3)** | **234,570 mm3** | **±83,998 mm3** | **[150,573 mm3, 318,572 mm3]** |
| **10mm  (524mm3)** | **20mm  (4,189 mm3)** | **3,665 mm3** | **±4145 mm3** | **[-480 mm3, 7,810 mm3]** |

**computed as**

**This claim holds when:**

* **the tumor is measurable at both timepoints (i.e., tumor margins are sufficiently conspicuous and geometrically simple enough to be recognized on all images in both scans; the tumor is unattached to other structures of equal density)**
* **the tumor longest in-plane diameter is between 10 mm (volume 0.5 cm3) and 100 mm (volume 524 cm3) at both timepoints**

Discussion

The **±**24% boundaries in Claim 1 can be thought of as “error bars” or “noise” around the measurement of volume change. If you measure change within this range, you cannot be certain that there has really been a change. However, if a tumor changes size beyond these limits, you can be 95% confident there has been a true change in the size of the tumor, and the perceived change is not just measurement variability. Note that this does not address the biological significance of the change, just the likelihood that the measured change is real.

Clinical interpretation (progression/response):   
To be 95% confident there has been a true increase or decrease in tumor volume, the measured change should be at least 24% for a tumor that had a longest in-plane diameter of between 50mm and 100mm at baseline (and at least 29% or 39% for the next two size categories respectively).

Alternatively, if the 95% confidence interval ported by the Analysis Software for the measured volume change (see 3.6.2) does not contain zero, it is highly likely there has been a true increase or decrease in tumor volume.

Whether such a true increase or decrease in tumor volume constitutes clinically meaningful disease progression or response is a distinct decision that requires a clinician’s judgment. There are currently no validated response criteria based on volume. The most commonly used response criteria in solid tumors, RECIST 1.1, uses unidimensional measurements. For comparison, RECIST 1.1 specifies that progression has occurred when there has been a 20% increase in tumor diameter, which corresponds to a 73% increase in volume for a spherical tumor, and favorable treatment response has occurred when there has been a 30% decrease in diameter, which corresponds to a 66% decrease in volume.

Clinical interpretation (magnitude of change):   
The magnitude of the true change is described in Claim 4 in terms of the 95% Confidence Interval of the measured volume change value. If you measure the volume to be 33,510mm3 at baseline and 268,080mm3 at follow-up (corresponding to a diameter change from 40mm to 80mm), then the 95% confidence interval for the true change is an increase in volume of 234,570 mm3 ±83,998.

The lower bound of 10mm on the tumor longest in-plane diameter is set to limit the variability introduced when approaching the resolution of the dataset, e.g. partial volume. The upper bound of 100mm is set to limit the variability introduced by more complex tumor morphology and organ involvement, and also to keep performance assessment procedures manageable.

While the claims have been informed by an extensive review of the literature and expert consensus that has not yet been fully substantiated by studies that strictly conform to the specifications given here. The expectation is that during field test, data on the actual field performance will be collected and any appropriate changes made to the claim or the details of the Profile. At that point, this caveat may be removed or re-stated.

The performance values in Claims 1, 2 and 3 reflect the likely impact of variations permitted by this Profile. The Profile requires that for a given tumor the same compliant radiologist actor and image analysis tool actor must make the measurement at both timepoints. If a different radiologist and/or image analysis tool was used at the baseline, this means the current radiologist and image analysis tool must repeat the baseline measurement for the result to be conformant with this profile. The profile permits the other actors (acquisition device, technologist, physicist, etc.) to differ at the two timepoints (i.e. it is not required that the same scanner be used for both exams of a patient). If one or more of the actors that are permitted to differ are the same, the implementation is still compliant with this Profile and it is expected that the measurement performance will be improved. To give a sense of the possible improvement, the following table presents expected precision for alternate scenarios, however except for the bolded column, these precision values are **not** Claims of this Profile. If the radiologist or image analysis tool are different (or any other requirement of the profile is not met), the measurement might still be clinically useful, but the measurement is no longer conformant with the Profile and the measurement claims should not be presumed.

Table 1: Minimum Detectable Differences for Tumor Volume Changes (Informative)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tumor Diameter** | **Different**  **Acquisition Device** | | | | Same  Acquisition Device | | | |
| Different  Radiologist | | **Same**  **Radiologist** | | Different  Radiologist | | Same  Radiologist | |
| Different Analysis Tool | Same Analysis Tool | Different Analysis Tool | **Same Analysis Tool** | Different Analysis Tool | Same Analysis Tool | Different Analysis Tool | Same Analysis Tool |
| >50mm | 43% | 24% | 43% | **24%** | 37% | 10% | 37% | 8% |
| 35-49mm | 67% | 33% | 65% | **29%** | 62% | 22% | 60% | 14% |
| 10-34mm | 139% | 120% | 80% | **39%** | 136% | 117% | 75% | 28% |

Notes:

1. Acquisition Device actors being different means the scanner used at the two timepoints were different models (from the same or different vendors). Two scanners with different serial numbers but of the same model are considered to be the same Acquisition Device actor.

2. Precision is expressed here as the total deviation index.

3. A measured change in tumor volume that exceeds the relevant precision value in the table indicates 95% confidence in the presence of a true change.

4. Minimum detectable differences calculated from the following formula: 1.96 x sqrt [2 x (wCV2 + bias2)], where wCV is estimated from the square root of the sum of the variances from the applicable sources of uncertainty (which makes the assumption that the variance components are additive, an assumption that has not yet been tested).

5. The numbers in the table combine factors such as the expected bias and wCV. The estimates of these values were derived from several groundwork studies, some of which were performed on phantoms and some of which were performed on human subjects.

# 3. Profile Requirements

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.

Conformant Actors shall support the listed Activities by conforming to all requirements in the referenced Section.

Table 1: Actors and Required Activities

|  |  |  |
| --- | --- | --- |
| **Actor** | **Activity** | **Section** |
| Acquisition Device | Periodic QA | 3.1. |
| Subject Handling | 3.2. |
| Image Data Acquisition | 3.3. |
| Physicist | Periodic QA | 3.1. |
| Image Data Reconstruction | 3.4 |
| Technologist | Subject Handling | 3.2. |
| Image Data Acquisition | 3.3. |
| Image Data Reconstruction | 3.4. |
| Radiologist | Subject Handling | 3.2. |
| Image Data Acquisition | 3.3 |
| Image Data Reconstruction | 3.4 |
| Image QA | 3.5. |
| Image Analysis | 3.6. |
| Reconstruction Software | Image Data Reconstruction | 3.4. |
| Image Analysis Tool | Image Analysis | 3.6. |

The requirements in this Profile do not codify a Standard of Care; they only provide guidance intended to achieve the stated Claim. Failing to conform to a “shall” in this Profile is a protocol deviation. Although deviations invalidate the Profile Claim, such deviations may be reasonable and unavoidable and 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.

For the Acquisition Device, Reconstruction Software and Image Analysis Tool actors, while it will typically be the manufacturer who claims the actor is conformant, it is certainly possible for a site to run the necessary tests/checks to confirm compliance and make a corresponding claim. This might happen in the case of an older model device which the manufacturer is no longer promoting, but which a site needs a compliance claim to participate in a clinical trial.

The Physicist actor represents the person at the site responsible for managing the equipment performance related specifications. At some sites this will be a staff physicist, and at other sites it may be a person who manages a contractor or a service provided by a vendor.

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

Acq.

*Subtract*

*volumes*

Subject

Handling

Recon



*Obtain images per timepoint (2)*

*Imaging*

*Agent*

(

*if any*

)

images

*Measure change per target lesion*

*Measure change in target lesion volume*

*Calculate*

*volume*

*Calculate*

*volume*

volume

changes

*volumes*

*...*

QA

Image Analysis

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 evaluated. 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 tumor by calculating absolute volume at each time point and subtracting. When expressed as a percentage, volume change is the difference in volume between the two time points divided by the volume at time point 1. Although this introduces some asymmetry (volume measurements of 50cm3 and 100cm3 represent either a 100% increase or a 50% decrease depending on which was measured first), it is more familiar to clinicians than using the average of the two timepoints as the denominator.

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 tumors as target is also beyond the scope of this document.

The Profile does not intend to discourage innovation, although it strives to ensure that methods permitted by the profile requirements will result in performance that meets the Profile Claim. The above pipeline provides a reference model. Algorithms which achieve the same result as the reference model but use different methods may be permitted, for example by directly measuring the change between two image sets rather than measuring the absolute volumes separately. Developers of such algorithms are encouraged to work with the appropriate QIBA committee to conduct any groundwork and assessment procedure revisions needed to demonstrate the requisite performance.

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.

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. Periodic QA

This activity describes calibrations, phantom imaging, performance assessments or validations performed periodically at the site, but not directly associated with a specific subject, that are necessary to reliably meet the Profile Claim.

### 3.1.1 Discussion

This activity is focused on ensuring that the acquisition device is aligned/calibrated/functioning normally. Performance measurements of specific protocols are not addressed here. Those are included in section 3.4.

### 3.1.2 Specification

| **Parameter** | **Actor** | **Requirement** |
| --- | --- | --- |
| QC | Physicist | Shall perform relevant quality control procedures as recommended by the manufacturer.  Shall record the date/time of QC procedures for auditing. |

## 3.2. Subject Handling

This activity describes details of handling imaging subjects that are necessary to reliably meet the Profile Claim.

### 3.2.1 Discussion

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.

**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.

**Timing Relative to Confounding Activities**

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

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

**Contrast Preparation and Administration**

Contrast characteristics influence the appearance, conspicuity, and quantification of tumor volumes.   
Non-contrast CT might not permit an accurate characterization of the malignant visceral/nodal/soft-tissue tumors and assessment of tumor boundaries.

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

It is important that the **Contrast Protocol** achieves a consistent phase and degree of enhancement. Bolus tracking is a good tool if available, but is not required. When using bolus tracking, be consistent between scans with where the ROI used for triggering is placed and the threshold used to trigger the scan. When bolus tracking is not available, be consistent between the scans with the contrast volume, rate, scan timing after injection, and use (or lack) of a saline flush. 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.

**Subject Positioning**

Positioning the subject Supine/Arms Up/Feet First has the advantage of promoting consistency (if it’s always the same, then it’s always consistent with baseline), and reducing cases where intravenous lines go through the gantry, which could introduce artifacts. 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.

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.

Artifact sources, in particular metal and other high density materials, can degrade the reconstructed volume data such that it is difficult to determine the true boundary of a tumor. Due to the various scan geometries, artifacts can be induced some distance from the artifact source. The simplest way to ensure no degradation of the volume data is to remove the artifact sources completely from the patient during the scan, if feasible. Although artifacts from residual oral contrast in the esophagus could affect the measurement of small tumors near the esophagus, this is not addressed here.

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

**Instructions to Subject During Acquisition**

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.

**Timing/Triggers**

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

### 3.2.2 Specification

| **Parameter** | **Actor** | **Specification** |
| --- | --- | --- |
| Contrast Protocol | Radiologist | Shall prescribe a contrast protocol that achieves enhancement consistent with baseline. |
| Use of intravenous contrast | Radiologist | Shall determine whether the selected contrast protocol, if any, will achieve sufficient tumor conspicuity. |
| Technologist | Shall use the prescribed intravenous contrast parameters.  Shall document the total volume of contrast administered, the concentration, the injection rate, and whether a saline flush was used. |
| Use of oral contrast | Radiologist | Shall determine whether the selected contrast protocol, if any, will achieve sufficient tumor conspicuity |
| Technologist | Shall use the prescribed oral contrast parameters.  Shall document the total volume of contrast administered and the type of contrast. |
| Subject Positioning | 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. |
| Artifact Sources | Technologist | Shall remove or position potential sources of artifacts (specifically including breast shields, metal-containing clothing, EKG leads and other metal equipment) such that they will not degrade the reconstructed CT volumes. |
| Table Height & Centering | Technologist | Shall adjust the table height for the mid-axillary plane to pass through the isocenter.  Shall position the patient such that the “sagittal laser line” lies along the sternum (e.g. from the suprasternal notch to the xiphoid process). |
| Breath hold | 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.  Shall ensure that for each tumor the breath hold state is consistent with baseline. |
| Image Header | 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.). |
| Acquisition Device | Shall provide corresponding data entry fields. |
| Contrast-based Acquisition Timing | 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 (i.e. obtained in the same phase; arterial, venous, or delayed).  Shall ensure that the time-interval between the administration of oral contrast and the start of the image acquisition is consistent with baseline. (Note that the tolerances for oral timing are larger than for intravenous). |
| Image Header | Acquisition Device | Shall record actual timing and triggers in the image header by including the Contrast/Bolus Agent Sequence (0018,0012). |

## 3.3. Image Data Acquisition

This activity describes details of the data acquisition process that are necessary to reliably meet the Profile Claim.

### 3.3.1 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 tumors. 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 beyond requiring that certain fields be populated in the image header. 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.

**Image Header** recordings of the key parameter values facilitate meeting and confirming the requirements to be consistent with the baseline scan.

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 duration** 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.

Recording of Anatomic Coverage by the Acquisition Device may or may not depend on attention and interaction by the Technologist.

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

### 3.3.2 Specification

| **Parameter** | **Actor** | **Specification** | **DICOM Tag** |
| --- | --- | --- | --- |
| Acquisition Protocol | Radiologist | Shall prepare a protocol to meet the specifications in this table.  Shall ensure technologists have been trained on the requirements of this profile. |  |
| Technologist | Shall select a protocol that has been prepared for this purpose.  Shall report if any parameters are modified beyond these specifications. |  |
| Acquisition Device | Shall be capable of storing the protocol and performing scans with all the parameters set as specified in this table. |  |
| Total Collimation Width | Radiologist | Shall set to Greater than or equal to 16mm. | Total Collimation Width  (0018,9307) |
| IEC Pitch | Radiologist | Shall set to Less than 1.5. | Spiral Pitch Factor  (0018,9311) |
| Nominal Tomographic Section Thickness (T) | Radiologist | Shall set to Less than or equal to 1.5mm. | Single Collimation Width  (0018,9306) |
| Scan Plane (Image Orientation) | Technologist | Shall set Consistent with baseline. | Gantry/Detector Tilt (0018,1120) |
| Tube Potential (kVp) | Technologist | Shall set Consistent with baseline (i.e. the same kVp setting if available, otherwise as similar as possible). | KVP  (0018,0060) |
| Scanogram | Technologist | Shall confirm on the scanogram the absence of artifact sources that could affect the planned volume acquisitions. |  |
| Scan Duration for Thorax | Technologist | Shall 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 | Technologist | Shall ensure the tumors to be measured and additional required anatomic regions are fully covered.  Shall, if multiple breath-holds are required, 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) |
| Acquisition Field of View (FOV) | Technologist | Shall set Consistent with baseline. | Data Collection Diameter (0018,0090) |
| Image Header | Acquisition Device | Shall record in the DICOM image header the actual values for the tags listed above in the DICOM Tag column. |  |

## 3.4. Image Data Reconstruction

This activity describes criteria and procedures related to producing images from the acquired data that are necessary to reliably meet the Profile Claim.

### 3.4.1 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 tumors. 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 beyond requiring that certain fields be populated in the image header. 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.

**Image Header** recordings of the key parameter values facilitate meeting and confirming the requirements to be consistent with the baseline scan.

**Spatial Resolution** quantifies the ability to resolve spatial details and scales the impact of partial volume effects. 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. If the spatial resolution is significantly different between the two timepoints, these impacts will change which can affect repeatability. So both balance and consistency is desirable. 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 assessed (See 4.1) in terms of the f50 value of the modulation transfer function (MTF) measured in a scan of a resolution phantom (such as module 1 of the CTAP phantom from the American College of Radiology). An implication of using the ACR phantom is that the resolution is assessed at only one distance from the isocenter. Although spatial resolution may vary with distance from the isocenter and tumors can be expected at various distances from the isocenter, it is considered fair to assume that resolution does not degrade drastically relative to the acceptable range of the resolution specification here.

Note that the noise and resolution specifications (See 3.4.2) to “ensure that the protocol in use has been validated in phantoms“ are not asking the tech to scan phantoms before every patient, or to validate the protocol themselves, just that the site needs to have validated the protocols that the tech will be using and conformance with the protocol depends on the tech selecting those protocols.

**Voxel 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. If algorithms become uniformly more "noise tolerant", the maximum threshold may be raised. Decreased image noise is not always beneficial, if achieved through undesirable image manipulation (e.g. extreme amounts of image smoothing), or scanning technique (e.g. increases in radiation dose or decreases in resolution). The profile does not currently define a minimum threshold, although it could be introduced as a means of forcing a balance between the goal of noise reduction, and other priorities.

The preferred metric for voxel noise is the standard deviation of reconstructed CT numbers over a uniform region in a phantom. The use of standard deviation 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.

Voxel noise (pixel standard deviation in a region of interest) can be reduced by reconstructing images with greater thickness for a given mAs. 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.

**Reconstructed Image Thickness** is the nominal width of the reconstructed image along the z-axis (reconstructed image thickness) since the thickness is not technically the same at the middle and at the edges.

**Reconstructed Image Interval** is the distance between two consecutive reconstructed images. An interval 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.

**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 **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.

### 3.4.2 Specification

**Note:** The Radiologist is responsible for the protocol parameters and validation, although they may be executed by the Physicist actor. The role of the Physicist actor may be played by an in-house medical physicist, a physics consultant or other staff (such as vendor service or specialists) qualified to perform the validations described.

| **Parameter** | **Actor** | **Specification** | **DICOM Tag** |
| --- | --- | --- | --- |
| Reconstruction Protocol | Radiologist | Shall prepare a protocol to meet the specifications in this table.  Shall ensure technologists have been trained on the requirements of this profile. |  |
| Physicist | Shall validate the protocol as described below. |  |
| Technologist | Shall select a protocol that has been prepared and validated for this purpose.  Shall report if any parameters are modified beyond these specifications. |  |
| Reconstruction Software | Shall be capable of performing reconstructions and producing images with all the parameters set as specified in this table. |  |
| Reconstructed Image Thickness | Radiologist | Shall set to between 1.0mm and 2.5mm (inclusive) and consistent (i.e. within 0.5mm) with baseline. | Slice Thickness (0018,0050) |
| Reconstructed Image Interval | Radiologist | Shall set to less than or equal to the Reconstructed Image Thickness (i.e. no gap, may have overlap) and consistent with baseline. | Spacing Between Slices (0018,0088) |
| Reconstruction Characteristics | Radiologist | Shall set the kernel and parameters Consistent with baseline (i.e. the same kernel and parameters if available, otherwise the kernel most closely matching the kernel response of the baseline). | Convolution Kernel Group (0018,9316), Convolution Kernel (0018,1210) |
| In-plane Spatial Resolution | Physicist | Shall validate that the protocol achieves an f50 value that is between 0.3 mm-1 and 0.7 mm-1, and is within 0.2 mm-1 of the baseline.  See 4.1. Assessment Procedure: In-plane Spatial Resolution |  |
| Voxel Noise | Physicist | Shall validate that the protocol achieves:   * a standard deviation that is < 50HU and consistent with the protocol used for the baseline scan within 5HU.   See 4.2. Assessment Procedure: Voxel Noise |  |
| Reconstruction Field of View | Technologist | Shall ensure the Field of View spans at least the full extent of the thoracic and abdominal cavity, but not substantially greater than that, and is consistent with baseline. | Reconstruction Field of View (0018,9317) |
| Image Header | Reconstruction Software | Shall record in the DICOM image header the actual values for the tags listed above in the DICOM Tag column as well as the model-specific Reconstruction Software parameters utilized to achieve compliance. |  |

## 3.5. Image QA

This activity describes criteria and evaluations of the images that are necessary to reliably meet the Profile Claim.

### 3.5.1 Discussion

This Image QA activity represents the portion of QA performed between image generation and analysis where characteristics of the content of the image are checked for compliance with the profile. The Image QA details listed here are the ones QIBA has chosen to highlight in relation to achieving the Profile claim. It is expected that sites will perform many other QA procedures as part of good imaging practices.

The Radiologist is identified here as ultimately responsible for this activity; however sites may find it beneficial for technologists to review these details at the time of imaging and identify cases which might require repeating acquisition and/or reconstruction to address issues with patient motion or artifacts.

Similarly, some or all of these checks may be performed at reporting time and as a result some or all of the tumor measurements may then be identified as not falling within the performance Claim of the Profile.

**Patient motion artifacts** can manifest in a variety of ways, such as a perceptible tram tracking appearance of the bronchioles or blurring of the lung architectural contours with lung windows.

**Dense object artifacts** (both internal and external to the patient) can variably degrade the ability to assess tumor boundaries as discussed in 3.2, resulting in poor change measures and repeatability.

**Clinical conditions** can also degrade the ability to assess tumor boundaries, or influence the structure of the tumor itself. For example, atelectasis, pleural effusion, pneumonia and/or pneumothorax can result in architectural changes to the lung surrounding a nodule. Necrosis may complicate decisions on the tumor extent.

**Tumor Size** can affect the accuracy of measurements. Both theoretical considerations and the groundwork projects done by QIBA indicate that for tumors that are small, errors in measurement represent a greater percentage of the measured size. For tumors that are smaller than the limits defined in this profile, please see the profile produced by the QIBA Small Nodule group for more information on imaging recommendations and performance claims. For tumors that are extremely large, the limitations on measurement are based less on imaging physics and more on anatomy. Such tumors are likely to cross anatomical boundaries and abut structures that make consistent segmentation difficult.

**Tumor Margin Conspicuity** refers to the clarity with which the boundary of the tumor can be discerned from the surroundings. Conspicuity can directly impact the ability to segment the tumor to properly determine its volume. Conspicuity problems can derive from poor contrast enhancement, from the inherent texture, homogeneity or structure of the tumor, or from attachment of the tumor to other structures.

**Tumor Measurability** is a general evaluation that is essentially left to the judgement of the radiologist, and it is their responsibility to oversee segmentation and disqualify tumors with poor measurability or inconsistent segmentation between the two timepoints. If the tumor has varying margin conspicuity on different slices, or is conspicuous but has complex geometry, or the segmentation software is visibly failing, or the background didn't respond to contrast the same way in the two time points, the radiologist should disqualify the tumor. Conversely, if the tumor is attached to another structure but the radiologist is confident they can get consistent segmentation over the two timepoints, they may allow a tumor that would be otherwise disqualified.

**Tumor Shape** is not explicitly identified as a specification parameter. No specific tumor shapes are considered a priori unsuitable for measurement. Although groundwork has shown that consistent measurements are more readily achieved with simple shapes than with complex shapes (such as spiculated tumors), the parameters for tumor size, tumor margin conspicuity and tumor measurability are felt to be sufficient. Moreover, complex shapes are even more difficult to assess accurately using simple linear measurements, increasing the relative added value of volumetry.

Keep in mind that this Profile is “lesion-oriented”. If one tumor in a study is excluded from the Profile Claim because the tumor does not comply with the specifications in this section, that does not affect other tumors in the same study which do comply with these specifications at both time points. Further, if a future study results in the excluded tumor being compliant at two time points, then the claim holds across those two time points.

While the radiologist is responsible for confirming case compliance with the Image QA specifications in Section 3.5.2, it is left to individual sites to determine the best approach in their work environment for capturing this audit data. Possible approaches include the use of a QIBA worksheet that captures this information, or asking the radiologist to dictate each parameter into the clinical report (e.g. the scan is free of motion or dense object artifacts, contrast enhancement is consistent with baseline, the tumor margins are sufficiently conspicuous").

### 3.5.2 Specification

The Radiologist shall ensure that the following specifications have been evaluated for each tumor being measured.

| **Parameter** | **Actor** | **Specification** |
| --- | --- | --- |
| Patient Motion Artifacts | Radiologist | Shall confirm the images containing the tumor are free from artifact due to patient motion. |
| Dense Object Artifacts | Radiologist | Shall confirm the images containing the tumor are free from artifact due to dense objects, materials or anatomic positioning. |
| Clinical Conditions | Radiologist | Shall confirm that there are no clinical conditions affecting the measurability of the tumor. |
| Tumor Size | Radiologist | Shall confirm (now or during measurement) that tumor longest in-plane diameter is between 10 mm and 100 mm.  (For a spherical tumor this would roughly correspond to a volume between 0.5 cm3 and 524 cm3.) |
| Tumor Margin Conspicuity | Radiologist | Shall confirm the tumor margins are sufficiently conspicuous and unattached to other structures of equal density to distinguish the volume of the tumor. |
| Contrast Enhancement | Radiologist | Shall confirm that the phase of enhancement and degree of enhancement of appropriate reference structures (vascular or tissue) are consistent with baseline. |
| Tumor Measurability | Radiologist | Shall disqualify any tumor they feel might reasonably degrade the consistency and accuracy of the measurement.  Conversely, if artifacts or attachments are present but the radiologist is confident and prepared to edit the contour to eliminate the impact, then the tumor need not be judged non-compliant with the Profile. |
| Consistency with Baseline | Radiologist | Shall confirm that the tumor is similar in both timepoints in terms of all the above parameters. |

## 3.6. Image Analysis

### 3.6.1 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 Accuracy** can affect the variability of Tumor Volume Change results. The volume accuracy is assessed to confirm that volume is being computed correctly and confirm there is a reasonable lack of bias at individual timepoints.

**Tumor Volume Change Repeatability**, 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 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%.

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.

**Reading Paradigms** (such as the “sequential locked” paradigm described here) can reduce variability from inconsistent judgments (such as where to separate an attached tumor) but also have the potential to introduce subconscious biases. The current edition of the profile does not prohibit the Image Analysis Tool from displaying the actual volume value from the previous timepoint since that might unnecessarily disqualify existing products. If it is determined to be the source of problems, it might be prohibited in future editions. Also, note that while the Image Analysis Tool is required to be capable of displaying the image from the previous timepoint, the radiologist is not required to look at it for every case. It is up to their judgment when to use that capability.

The profile requires that the same Image Analysis Tool and the same Radiologist perform the measurement of both timepoints of a given tumor. This requirement is due to the variability introduced when a different Image Analysis Tool and/or Radiologist is used between the two timepoints. See Table 1 and the related Claim Discussion for more details.

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.

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 manufacturer 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 manufacturers 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.

**Confidence Interval of Result** provides a range of plausible values for the change in tumor volume. The 95% confidence interval (CI) can be interpreted as follows: If the change in a tumor's volume over two timepoints is measured repeatedly and the 95% CI constructed for each measurement, then 95% of those CIs would contain the true volume of the tumor.

A reference implementation of a calculator that uses the specified equation is available at the following location: <http://www.accumetra.com/NoduleCalculator.html>

**Recording** various details can be helpful when auditing the performance of the biomarker and the site using it. For example, it is helpful for the system to record the set-up and configuration parameters used, or to be capable of recording the tumor segmentation as a DICOM Segmentation. Systems based on models should be capable of recording the model and parameters.

It is up to products that do not use contours to propose a method for verification by the radiologist.

### 3.6.2 Specification

| **Parameter** | **Actor** | **Specification** |
| --- | --- | --- |
| Multiple Tumors | Image Analysis Tool | Shall allow multiple tumors to be measured.  Shall either correlate each measured tumor across time points or support the radiologist to unambiguously correlate them. |
| Reading Paradigm | Image Analysis Tool | Shall be able to present the reader with both timepoints side-by-side for comparison when processing the second timepoint.  Shall re-process the first time point if it was processed by a different Image Analysis Tool or Radiologist. |
| Radiologist | Shall re-process the first time point if it was processed by a different Image Analysis Tool or Radiologist. |
| Tumor Volume Computation | Image Analysis Tool | Shall be validated to compute tumor volume with accuracy within 3 % of the true volume.  See 4.3 Assessment Procedure: Tumor Volume Computation. |
| Tumor Volume  Change Repeatability | Image Analysis Tool | Shall be validated to achieve tumor volume change repeatability with:   * an overall repeatability coefficient of less than 16%. * a small subgroup repeatability coefficient of less than 21% * a large subgroup repeatability coefficient of less than 21%   See 4.4. Assessment Procedure: Tumor Volume Change Repeatability. |
| Radiologist | Shall, if operator interaction is required by the Image Analysis Tool to perform measurements, be validated to achieve tumor volume change repeatability with:   * an overall repeatability coefficient of less than 16%. * a small subgroup repeatability coefficient of less than 21% * a large subgroup repeatability coefficient of less than 21%   See 4.4. Assessment Procedure: Tumor Volume Change Repeatability (Image Analysis Tool). |
| Tumor Volume Bias & Linearity | Image Analysis Tool | Shall be validated to achieve:   * an overall tumor volume %bias of less than shown in Table 3.6.2-2 (below) * a tumor volume %bias for each shape subgroup (spherical, ovoid, lobulated) of less than shown in Table 3.6.2-2 (below) * slope ( between 0.98 and 1.02   Values are taken from Table 3.6.2-2 based on the overall repeatability coefficient achieved by the Image Analysis Tool using the assessment procedure in 4.4.  See 4.5 Assessment Procedure: Tumor Volume Bias and Linearity. |
| Result  Verification | Radiologist | Shall review & approve margin contours produced by the tool. |
| Confidence Interval of Result | Image Analysis Tool | Shall calculate and make available to the operator the 95% confidence interval for tumor volume change based on the equation:  Where  *Y1* and *Y2* is the volume measurement at timepoint 1 and 2,  *wCV1* and *wCV2* is the within-nodule coefficient of variation  for *Y1* and *Y2* as taken from the following table,  *D1* and *D2* is the longest in-plane diameter of the volume at  timepoint 1 and 2:     |  |  |  |  | | --- | --- | --- | --- | | ***D1*, *D2*** | **10-34mm** | **35-49mm** | **50-100mm** | | ***wCV1*, *wCV2*** | 0.501 | 0.241 | 0.157 | |
| Result Recording | Image Analysis Tool | Shall record the percentage volume change relative to baseline for each tumor.  Shall record the confidence interval of result for each change measurement.  Shall record the image analysis tool version. |

**Table 3.6.2-2:   
Allowable Tumor Volume %Bias based on Repeatability Coefficient**

|  |  |  |
| --- | --- | --- |
| **Repeatability Coefficient p** | **Allowable Overall %Bias** | **Allowable Shape Subgroup %Bias** |
| 5% | <7.2% |  |
| 6% | <7.0% |  |
| 7% | <6.8% |  |
| 8% | <6.5% |  |
| 9% | <6.3% |  |
| 10% | <5.9% |  |
| 11% | <5.5% |  |
| 12% | <5.1% |  |
| 13% | <4.5% |  |
| 14% | <3.8% |  |
| 15% | <2.8% |  |
| 16% | <1.2% |  |

# 4. Assessment Procedures

To conform to this Profile, participating staff and equipment (“Actors”) shall support each activity assigned to them in Table 1.

To support an activity, the actor shall conform to the checklist of requirements (indicated by “shall language”) listed in the specifications table of that activity subsection in Section 3.

Although most of the requirements described in Section 3 can be assessed for conformance by direct observation, some of the performance-oriented requirements cannot, in which case the requirement will reference an assessment procedure in a subsection here in Section 4.

Formal claims of conformance by the organization responsible for an Actor shall be in the form of a published QIBA Conformance Statement. Manufacturers 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 conformance. Manufacturers shall also provide access or describe the characteristics of the test set used for compliance testing.

## 4.1. Assessment Procedure: In-plane Spatial Resolution

This procedure can be used by a manufacturer or an imaging site to assess the In-plane Spatial Resolution of reconstructed images. Resolution is assessed in terms of the f50 value (in mm-1) of the modulation transfer function (MTF). Loosely speaking, the MTF represents the blur of an infinitely small feature of interest, f50 represents the spatial frequency at which the contrast of the feature has decreased by 50%, and the inverse of the f50 value represents the size of a feature that would be degraded 50%. So for an f50 value of 0.4 mm-1, features that are 2.5mm (or smaller) would have their contrast degraded by 50% (or more).

The assessor shall first warm up the scanner’s x-ray tube and perform calibration scans (often called air-calibration scans) according to scanner manufacturer recommendations.

The assessor shall scan a spatial resolution phantom, such as the ACR CT Accreditation Program (CTAP) Phantom’s module 1, which has a series of HU-value cylindrical inserts including one with soft-tissue equivalence. The acquisition protocol and reconstruction parameters shall conform to this Profile (See Section 3.3.2 and 3.4.2). The same protocol and parameters shall be used when performing the assessments in 4.1 and 4.2. I.e., the noise level during resolution assessment should correspond to that measured during noise assessment.

The phantom shall be positioned with the center of the phantom at isocenter and properly aligned along the z-axis as described in the ACR CTAP documentation about alignment of the beads.

When the scan is performed, the assessor shall generate an MTF curve, measured as an average of the MTF in the x-y plane along the edge of a target soft-tissue equivalent insert using AAPM TG233 methodology as implemented in manufacturer analysis software, AAPM TG233 software or equivalent.

The assessor shall then determine and record the f50 value, defined as the spatial frequency (in mm-1 units) corresponding to 0.5 MTF on the MTF curve.

The procedure described above is provided as a reference method. This reference method and the method used by the scanner manufacturer for FDA submission of MTF values are accepted methods for this assessment procedure. Note that for iterative reconstruction, the manufacturer may have specific test methodologies appropriate for the given algorithm.

Sites may submit to QIBA a proposed alternative method and evidence that the results produced by the proposed method are equivalent to this reference method or to the manufacturer method. Upon review and approval by QIBA, the alternative method will also become an accepted assessment procedure in this Profile.

The test procedure described here may be applied to both conventional filtered backprojection reconstruction methods and iterative reconstruction methods.

## 4.2. Assessment Procedure: Voxel Noise

This procedure can be used by a manufacturer or an imaging site to assess the voxel noise of reconstructed images. Voxel noise is assessed in terms of the standard deviation of pixel values when imaging a material with uniform density.

The assessor shall first warm up the scanner’s x-ray tube and perform calibration scans (often called air-calibration scans) according to scanner manufacturer recommendations. The assessor shall then scan a phantom of uniform density, such as the ACR CT Accreditation Program (CTAP) Phantom’s module 3, which is a 20 cm diameter cylinder of water equivalent material. The phantom shall be placed at the isocenter of the scanner. The acquisition protocol and reconstruction parameters shall be compliant with this Profile (See Section 3.3.2 and 3.4.2). The same protocol and parameters shall be used when performing the assessments in 4.1 and 4.2.

When the scan is performed, the assessor shall select a single representative slice from the uniformity portion of the phantom.

An approximately circular region of interest (ROI) of at least 400 mm2 shall be placed near the center of the phantom. The assessor shall record the values reported for the ROI mean and standard deviation.

The procedure described above is provided as a reference method. Sites may submit to QIBA a proposed alternative method (such as using the water phantom portion of a manufacturer’s QA phantom) and evidence that the results produced by the proposed method are equivalent to this reference method or manufacturer methodology. Upon review and approval by QIBA, the alternative method will also become an accepted assessment procedure in this Profile.

The test procedure described here is intended to be a simple phantom measurement that sets a reasonable floor on the noise which is considered sufficient to avoid degrading segmentation performance. The procedure may be used for both conventional filtered backprojection and iterative reconstruction methods. It is noted that when characterizing reconstruction methods, voxel noise is a limited representation of image noise when noise texture is varied.

## 4.3. Assessment Procedure: Tumor Volume Computation

This procedure can be used by a manufacturer or an imaging site to assess whether an Image Analysis Tool computes the volume of a single tumor correctly. Accuracy is assessed in terms of the percentage error when segmenting and calculating the volume of a tumor with known truth.

The assessor shall obtain the test files in DICOM format from the QIDW. They can be found by searching for the CT volumetry digital reference object (DRO) DICOM image set. The test files represent a digital test object with z-axis resolution of 1.5mm. A test nodule with -10 HU radio-density is placed within a flat -1000 HU region of the phantom to make the segmentation intentionally easy since the test is not intended to stress the segmentation tool but to instead evaluate any bias in the volume computation after the lesion is segmented.

The assessor shall use the Image Analysis Tool to segment and calculate the volume of the single tumor present in the test images.

The assessor shall record the percentage difference between the reported volume and the true value. The true value is provided in the description of the test files on QIDW.

## 4.4. Assessment Procedure: Tumor Volume Change Repeatability

This procedure can be used by a manufacturer or an imaging site to assess the repeatability with which the volume of a single tumor is measured. Repeatability is assessed in terms of the repeatability coefficient when segmenting and calculating the volume of a tumor with known truth. The procedure assesses an Image Analysis Tool and a Radiologist operating the tool as a paired system.

The assessment procedure has the following steps:

* Obtain a designated test image set (see 4.4.1).
* Determine the volume change for designated tumors (see 4.4.2).
* Calculate statistical metrics of performance (see 4.4.3).

Note that tumor detection is not evaluated by this procedure since the locations of the target lesions are provided.

### 4.4.1 obtain test image set

The test image set consists of multiple target tumors in the lung in multiple subjects which is representative of the stated scope of the Profile.

The assessor shall obtain the test files in DICOM format.

Lung tumor data is obtained from the Cancer Imaging Archive by searching for the test-retest subset of the RIDER Lung CT Dataset at (https://public.cancerimagingarchive.net/ncia/login.jsf).

The test files represent 31 cases, with two time points per case, each with one target tumor to segment. The target tumor is identified in terms of its x/y/z coordinates in the dataset. The list of target tumors and coordinates are provided in file: (???)

Future editions of the Profile may address a larger number of body parts (e.g., metastases in the mediastinum, liver, adrenal glands, neck, retroperitoneum, pelvis, etc. described in Appendix B.3) by including such tumors in the test data, and may test boundary condition performance by including test data that is marginally conformant (e.g. maximum permitted slice thickness, maximum permitted noise, etc.) to confirm conformant performance is still achieved.

The target tumors have been selected to be measureable (as defined in the Profile) and have a range of volumes, shapes and types to be representative of the scope of the Profile.

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

If the algorithm has been developed using the specified test files, that is unfortunate and shall be reported by the assessor.

### 4.4.2 determine volume change

The assessor shall segment each target tumor at each timepoint as described in the Image Analysis Activity (See 3.6). The assessor is permitted to edit the tumor segmentation or seed point if that is part of the normal operation of the tool. If segmentation edits are performed, results shall be reported both with and without editing.

When evaluating an Image Analysis Tool, a single reader shall be used for this entire assessment procedure.

When evaluating a Radiologist, a single tool shall be used for this entire assessment procedure.

The assessor shall calculate the volume (Y) of each target tumor at time point 1 (denoted Y*i*1) and at time point 2 (Y*i*2) where *i* denotes the *i*-th target tumor.

The assessor shall calculate the resulting % volume change (d) for each target tumor as

.

### 4.4.3 calculate statistical metrics of performance

The assessor shall calculate the within-subject Coefficient of Variation (wCV), where N=31 and

The assessor shall estimate the Repeatability Coefficient (RC) as

The assessor shall convert the Repeatability Coefficient (RC) estimate to a percentage as

.

The assessor shall divide the target tumors into a small subgroup (containing the 15 target tumors with the smallest measured volumes) and a large subgroup (containing the 16 tumors with the largest measured volumes). The assessor shall repeat the above calculations on both subgroups to estimate a small subgroup repeatability coefficient and a large subgroup repeatability coefficient.

The assessor is recommended to also compute Bland-Altman plots of the volume estimates as part of the assessment record.

For further discussion/rationale, see Annex E.2 Considerations for Performance Assessment of Tumor Volume Change.

## 4.5. Assessment Procedure: Tumor Volume Bias and Linearity

This procedure can be used by a manufacturer or an imaging site to assess the bias and linearity with which the volume of a single tumor is measured. Bias is assessed in terms of the percentage population bias when segmenting and calculating the volume of a number of tumors with known truth. Linearity is assessed in terms of the slope of an OLS regression fit to the volume data.

### 4.5.1 obtain test image set

The test image set consists of scans from two different scanners of an anthropomorphic ("Lungman") phantom with multiple synthetic target tumors of different shapes and sizes in the lung.

The assessor shall obtain the test files in DICOM format.

The data is obtained from the Cancer Imaging Archive by searching for the <<Nick will clarify the location of the FDA Lungman N1 data>> subset of the RIDER Lung CT Dataset at (https://public.cancerimagingarchive.net/ncia/login.jsf).

The test files represent 3 repeated scans of the FDA Lungman N1 phantom on each of 2 CT scanners. The phantom contains 7 synthetic tumors, each with a different combination of shape and diameter (see Table 4.5.1-1). The target tumors are identified in terms of their x/y/z coordinates in each of the 6 scans. The list of target tumors and coordinates are provided in file: (???)

Table 4.5.1-1: Phantom Target Tumor Characteristics

|  |  |  |
| --- | --- | --- |
| **Shape** | **Nominal Diameter** | **Nominal Density** |
| Spherical | 10 mm  20 mm  40 mm | +100 HU |
| Ovoid | 10 mm  20 mm | +100 HU |
| Lobulated | 10 mm  20 mm | +100 HU |

The target tumors have been placed to be measureable (as defined in the Profile) and have a range of volumes and shapes to be representative of the scope of the Profile.

The test image set has been acquired according to the requirements of this Profile (e.g. patient handling, acquisition protocol, reconstruction). See Table 4.5.1-2.

Table 4.5.1-2: Test Image Set Acquisition and Reconstruction Parameters

|  |  |
| --- | --- |
| **Scanner** | **Key Parameters** |
| Philips 16  (Mx8000 IDT) | KVp: 120  Pitch: 1.2  Collimation: 16x1.5  Exposure: 100 mAs  Slice Thickness: 2 mm  Increment: 1 mm  Filter: Medium  Repeat Scans: 3 |
| Siemens 64 | KVp: 120  Pitch: 1.2  Collimation: 64x0.6  Exposure: 100 mAs  Slice Thickness: 1.5 mm  Increment: 1.5 mm  Filter: Medium  Repeat Scans: 3 |

### 4.5.2 determine volume change

The assessor shall segment each of 42 target tumors (7 tumors in 3 scans for each of 2 scanners) as described in the Image Analysis Activity (See 3.6).

The assessor is permitted to edit the tumor segmentation or seed point if that is part of the normal operation of the tool. If segmentation edits are performed, results shall be reported both with and without editing.

When evaluating an Image Analysis Tool, a single reader shall be used for this entire assessment procedure.

When evaluating a Radiologist, a single tool shall be used for this entire assessment procedure.

The assessor shall calculate the volume (Y) of each target tumor (denoted Y*i*) where *i* denotes the *i*-th target tumor.

### 4.5.3 calculate statistical metrics of performance

The natural log of the true volumes (Xi) of each target tumor are known and are provided in the dataset.

The assessor shall calculate the individual bias (*bi*) of the measurement of each target tumor as

The assessor shall estimate the population bias over the N target tumors as

The assessor shall convert to a percentage bias estimate as

=

The assessor shall fit an ordinal least squares (OLS) regression of the on and shall estimate the slope ).

The assessor shall divide the target tumors into three subgroups (containing the spherical, ovoid and lobulated target tumors respectively). The assessor shall repeat the percentage population bias calculation on each subgroup to estimate a spherical subgroup percentage bias, an ovoid subgroup percentage bias and a lobulated subgroup percentage bias.

The assessor is recommended to also plot the volume estimate ( versus ) and the OLS regression curve of the volume estimates as part of the assessment record.

## 4.X. Assessment Procedure: Image Acquisition Site Performance

Note: The procedure in this section is currently only a proposal.

A more detailed procedure 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.X.1).
* Generate a test image set (see 4.X.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 3.6.2.

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.X.1 Acquisition Validation

Review patient handling procedures for compliance with Section 3.2

Establish acquisition protocols and reconstruction settings on the Acquisition Device compliant with Section 3.3 and Section 3.4. If a QIBA Conformance Statement is available from the Acquisition Device manufacturer, it may provide parameters useful for this step.

Acquire images of a 20cm water phantom, reconstruct and confirm performance requirements in Section 3.4.2 are met.

Discussion:

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

### 4.X.2 Test Image Set

Locally acquire a test image set using the protocols established and tested in Section 4.X.1.

The test image set should conform to the characteristics described in Section 4.X.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.

# 

# 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. It will be removed during publication of the Technically Confirmed Draft.

|  |  |
| --- | --- |
| **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 tumor 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. See Section 4. |
| **10** | **Q. What is the basis of the specification of 15% for the variability in tumor 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. |
| **16** | **Q. Should the claim (and profile) reflect reproducibility (actors must be compliant but are allowed to be different) or repeatability (actors must be compliant and must be the same)?**  A. State claim for scanner/reader/analysis-SW all permitted to be different across timepoints.  This is most applicable to clinical practice. Although QIBA started by looking at Clinical Trials, it has really evolved to address Clinical Practice and that is more generally useful and practical.  Different scanners cannot be avoided. Theoretically, different readers/SW could be avoided by requiring re-read/re-analyze of prior timepoints if different, but practically speaking, routine practice will not accommodate re-reading.  Note that when actors are not different across timepoints you are still compliant with the profile and performance can be expected to improve. If we can provide informative material about the degree of improvement, that would be helpful for some users. If there is minimal additional load in terms of assessment procedures, we can also consider elevating such alternate scenario performance to be part of the claim too. |
| **17** | Should assessment procedures be "open book" or "closed book"?  A: "Open book" for now.  With “closed book” the correct answers are not available to the assessor. This depends on someone setting up infrastructure for manufacturers/sites to submit data and a system to calculate and return a “closed book” score. May consider this in the future if there is sufficient need/value. |

# 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 (CTVol) Technical Committee. The CTVol 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 committee 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

# 

## 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 3.3, 3.4 and 3.6.

Some specific acquisition parameters, reconstruction parameters and analysis software parameters that are expected to be compatible with meeting the Profile requirements may be found in QIBA Conformance Statements published by manufacturers and sites. 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.

Manufacturers 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 with published QIBA Conformance Statements are encouraged to consider those parameters for both simplicity and consistency. Sites using models without published QIBA Conformance Statements 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.

Parameter Derivation Procedure

This procedure can be used by a manufacturer or an imaging site to select an appropriate reconstruction kernel??? for an acquisition device model: and various parameter sets that might also be compliant (but it’s not the only way that’s acceptable to QIBA).

* Set the scanning field of view for the patient, the Kyoto Kagaku chest phantom. This setting is to be used to image the ACR phantom. *Special handling*: In scanning the ACR CT phantom, some manufacturers specify the use of a FOV appropriate to the ACR device (See: <http://www.acr.org/accreditation/computed/ct_faq.aspx#thirteen>). In this case, follow the manufacturer’s guidance for the ACR phantom. As an example, for the Aquilion16 scanner see the guidance in slide 11 of

<http://www.tams-media.com/tams2008/sales/faqs/CT_pdfs/ACR_Guide_Aquilion16.pdf>

* Set the beam voltage to 120 kVp
* Set the slice thickness to between 0.75 and 1.25 mm (depending on the available reconstructed slice thicknesses of the scanner)
* Set nominal beam collimation (NxT such as 16 x 0.5mm, or 128 x 0.6mm, 320 x 0.5 mm) rotation time and pitch such that scan can cover a 35 cm thorax in 15 seconds or less
* ITERATE (hopefully only a few times) on reconstruction kernels to meet spatial resolution spec.



Figure : Establishing spatial resolution

* ITERATE (again, hopefully just a few times) on mAs or effective mAs setting, given beam collimation, pitch and rotation time.



Figure : Establishing noise spec

* If the scanning FOV is to be changed for the scan of the lung phantom, reset the FOV accordingly and rescan the ACR phantom. Measure the quality parameters, the noise and resolution, with the changed settings.

The quality parameters are expected to change under a changed scanning field of view, as in the special handling. If this is the case, the comparability of the quality of the various scanners is lost. In the analysis of the quantitative measurements of nodule sizes, the data on the actual quality measures may prove to be useful in analyzing device differences.

## Appendix E: Metrology Definitions and Methods

Two statistical analyses were conducted, based on the type of data: 1) variability of scalar volume measurements, including individual participant performance across test-retest repetitions as well as the performance across algorithms, and 2) comparison of segmentation boundaries relative to reference standard segmentations. The former allows us to compare the performances of these imaging algorithms by measuring agreement of the computed result when the algorithm is held constant as well as when measured by different algorithms, regardless of the similarity in the contours that give rise to the scalar volumes; the latter provides the means by which differing algorithms may be evaluated in terms of the specific segmentation task they are performing which gives rise to the computed scalar volumes.

**Variability of Scalar Volume Measurements**

The models we use assume that the variance is constant across the range of the response. As a first step we determine whether the input data is skewed, which indicates a non-constant variance and perform a transformation suited to the distribution presented so as to achieve constant variance. For this data, the measurement variation was not constant across the range of volumes; it increased with increasing volume measurements as will be illustrated below. Typically, input data is log normal, hence the selection of log transformation as appropriate. To meet the assumptions of the analyses, a log-transformation was applied to volume. As a result, residuals approximately follow a normal distribution which validates the conclusions of the model outputs. Whereas analyses were conducted on the log-scale, data is presented on the original scale, where possible.

Based on the transformed data, we undertook two analyses of volume measurement variability in this study, *repeatability* and *reproducibility* [[22](#_ENREF_22)]. Repeatability refers to variability of measuring tumor volume when repeated measurements are acquired on the same subject under identical or nearly identical conditions. Thus assessment of repeatability approximates the “pure” measurement error of tumor volume measurement. Specifically, we assessed repeatability as the variance of tumor volume measurement when the marker was obtained from repeated imaging of subjects with intentionally short interval so that biological features could be reasonably assumed to have remained unchanged. We assessed repeatability for each of the several participating groups. We further assessed reproducibility, as the variability in tumor volume measurements under the condition where algorithms are not held constant.

We used visual as well as numeric methods to assess variability. Plotting test-retest replications (for repeatability) or pair-wise combinations of algorithms (reproducibility) appear as a straight line of unity in the presence of agreement. Numerically, we denote the measurement of the *j*th algorithm for the *i*th subject at the *kth* replication as *Yijk*, where *j*=1,…,11, *i*=1, …,31, and *k*=1, 2. We used a simple general model , where *Yijk* and *εijk* are the observed value and measurement error and where *μ* is the population mean. is conditional on the mean of infinite replications made on subject *i* by algorithm *j*. Both repeatability and reproducibility were assessed numerically as well as graphically by the Bland and Altman method [[45](#_ENREF_45), [46](#_ENREF_46)]. The method produces an Upper Agreement Limit (*UAL*) and the Lower Agreement Limit (*LAL*) which provides a range within which we expect 95% of the differences between replicate measures of a given algorithm (repeatability) or pair-wise measures of by two algorithms (reproducibility), are expected to lie. Repeatability was represented as differences between each test-retest repetition and plotted against the averages of the two volume measurements.

Based on these analyses, we compute multiple metrics because each provides complementary insight into performance. *RC* is the least significant difference between two repeated measurements on a case taken under the same conditions,

.

The interpretation of *RC* is that the difference between any two normally-distributed measurements on the subject is expected to fall between –*RC* and *RC* for 95% of replicated measurements [[47](#_ENREF_47)]. The 95% tolerance interval for 95% of differences between replicated measurements is , where

and is the thpercentile of the distribution with degrees of freedom.

The within-subject standard of deviation (*wSD*) is estimated as square root of the averaged sample variances across tumors, where the sample variance is computed from the replications for each tumor. This *wSD* assumes that the within-tumor variance is the same across all tumors. The within-subject coefficient of variance (*wCV*) is a relative measure of repeatability, which we calculate as *wSD*/mean and thus is proportional to the magnitude of the tumor’s size.

Concordance correlation coefficient (*CCC)* was computed as in [[48](#_ENREF_48)]. *CCC* is a measure of agreement that is a product of the correlation coefficient, penalized by a bias term that reflects the degree to which the regression line differs from the line of agreement.

Reproducibility was analyzed similarly but instead of the two repetitions, pairwise comparisons were made between algorithms. In this case, the LOA by Bland and Altman provides a range within which we expect 95% of the differences in measurements between two algorithms to lie. The LOA are calculated as ± t(n-1); α/2 *sd*(*Yi1k* – *Yi2k*) (1+1/n), where , where *i*, *j*, *k*, t, and n are as defined above in the repeatability section but now *j* varies pair-wise. Linear Mixed Effects (LME) modeling is used to separate the variability due to subject, algorithm, subject-by-algorithm interaction, and residual. Each of these terms was considered as a random effect in the model. Model assumptions were evaluated with a Q-Q and observed-versus-fitted plots. Use of this model determined the relative contribution to variability by the algorithm as assessed as the sum of the variability to algorithm summed with the subject-by-algorithm interaction as compared with the residual due to other factors in order to inform the QIBA claim by measuring to what extent algorithm versus other variance contributes to overall error.

As done with repeatability, we compute multiple metrics from the reproducibility analysis because each provides complementary insight into performance. The reproducibility coefficient (*RDC*), the interclass correlation coefficient (*ICC*), and the magnitude of variance explained by algorithm versus the residual variance which originates in other factors. Similar to *RC*, *RDC* is calculated as the least significant difference between two measurements taken under different conditions, in our project, by two different algorithms. The *ICC* is a measure of the agreement between the participating groups’ measurements of the CT volumes. The *ICC* is a relative index; it depends on the between-tumor variability. Since the between-tumor variability differs in magnitude for small and large tumors, the *ICC*s of the small and large tumors are not comparable.

In addition to computing the metrics on all tumors, two stratified reproducibility analyses were performed, one by the degree of automation used by the algorithm, and a second by retrospectively dividing into two types: (a) tumors that could be classified as meeting the conditions described in the “Claims” section of the QIBA Profile [[43](#_ENREF_43)], and (b) tumors that did not meet these conditions. Specifically, the claims section of the QIBA profile states that the claims are only applicable “*when the given tumor is measurable (i.e. tumor margins are sufficiently conspicuous and geometrically simple enough to be recognized on all images….) .and the longest in-plane diameter of the tumor is 10 mm or greater*”. Therefore, tumors described as meeting the QIBA Profile were those that were judged to have clearly identified tumor margins; all tumors used in this study exceeded the 10 mm diameter threshold.

**Comparison of Segmentation Boundaries**

Whereas the nature of clinical data makes actual ground truth unavailable, we can form a reference truth if one assumes that those pixels with the highest agreement among participants as being part of the tumor (or not part of the tumor), may collectively be said to be a reference segmentation. We first produced a reference segmentation using the Simultaneous Truth And Performance Level Estimation (STAPLE) method [[49](#_ENREF_49)]. This filter performs a pixel-wise combination of an arbitrary number of input images, where each of them represents a segmentation of the same image, i.e., the segmentations performed by participant algorithms. Each input segmentation is weighted based on its "performance" as estimated by an expectation-maximization algorithm, described in detail in [[50](#_ENREF_50)]. We then compare each individual segmentation result to this reference data, using Sensitivity (*SE*) or true positive rate, calculated as follows. If we define a confusion matrix *C* where *Cuv* is the number of voxels segmented with label *u* while the true label is *v*. For any label *w*, we calculate true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*) as:

Typically *SE* is accompanied by Specificity, otherwise known as the true negative rate. However, this quantity has a strong dependence on the size of the field of view which is constant for all participants so we omit reporting this as it is not informative. Rather, the otherwise unused *TP* and *FN* computations are used in the calculation of two additional spatial overlap measures, the Jaccard index [[51](#_ENREF_51)], and Sørensen–Dice coefficients [[52](#_ENREF_52), [53](#_ENREF_53)]:

While at some point it may be evident which is the more important, for this work we compute and present all three types of numeric comparisons, collectively described as “overlap metrics.”