QIBA Profile:

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

Stage: Consensus+ Draft Tech Confirmed

When referencing this document, please use the following format:

QIBA CT Volumetry Technical Committee. CT Tumor Volume Change Profile - 2016, Consensus Profile. Quantitative Imaging Biomarkers Alliance, November 21, 2016. Available at: http://qibawiki.rsna.org/index.php/Profileshttp://rsna.org/QIBA_.aspx

Table of Contents

1. Executive Summary

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

Profile development is an evolutionary, phased process; this Profile is in the Consensus stage. The performance claims represent expert consensus and will be empirically demonstrated at a subsequent stage. Users of this Profile are encouraged to refer to the following site to understand the document's context: http://qibawiki.rsna.org/index.php/QIBA_Profile_Stages.

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 for Advanced Disease) 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 biomarker performance target is that:

> A true change in a tumor volume has occurred with 95% confidence if the measured change is larger than 24%, 29% or 39% when the longest in-plane diameter is initially 50-100mm, 35-49mm or 10-34mm, respectively.

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.

For convenience, the QIBA website also provides a "checklist" document which has re-grouped the requirements from Section 3 for each Actor to more easily communicate and confirm conformance of sites, staff and equipment to this Profile.

Note that this Profile document only states requirements to achieve the claim, not "requirements on standard of care."
Further, 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 thoracic tumors and measuring tumor longitudinal changes within subjects (i.e. evaluating growth or regression with image processing of CT scans acquired at different timepoints).

Conformance with this Profile by all relevant staff and equipment supports the following claims (see Disclaimer in Discussion below):

Claim 1: A true change in a tumor volume has occurred with 95% confidence if the measured volume change is larger than 24% ~~and~~ when 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 volume change is larger than 29% ~~and~~ when 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 volume change is larger than 39% ~~and~~ when the longest in-plane diameter is initially 10-34mm.

Claim 4: The tumor volume measurement performance, expressed as within-tumor coefficient of variation (wCV), is 0.085, 0.103, and 0.141 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 Confidence Interval Calculation	95% Confidence Interval of True Volume Change
100mm (524 cm^3)	50mm (65 cm^3)	-459 cm^3 ± 88 cm^3	[-547 cm^3, -371 cm^3]
40mm (34 cm^3)	80mm (268 cm^3)	234 cm^3 ± 45 cm^3	[189 cm^3, 279 cm^3]
10mm (0.5 cm^3)	20mm (4.2 cm^3)	3.7 cm^3 ± 1.2 cm^3	[2.5 cm^3, 4.9 cm^3]

computed as $(Y_2 - Y_1) \pm 1.96 \times \sqrt{(Y_1 \times wCV_1)^2 + (Y_2 \times wCV_2)^2}$, where Y_1 and Y_2 are the volume measurements at baseline and the subsequent timepoint, and wCV_1 and wCV_2 are the wCV estimates corresponding to these measurements.

These claims hold 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 cm^3) and 100 mm (volume 524 cm^3) at both timepoints**

Commented [OK1]: Flip to volume (diameter). As you move across the row, it disoriented people. The measurement is all about volume. The diameter is just for binning.

Commented [OK2]: Add a note explaining why a 524 cm3 tumor can shrink by 547 cm3. (since the original 524 measurement might be off)

Discussion

Disclaimer: While this profile is written to be applicable to thoracic tumors, the quantitative performance values were derived from analysis of tumor volumetry consisting solely of lung data. The claims assert that this performance holds for tumors throughout the thorax based on the expert opinion of key contributors to this profile who anticipate that performance for segmentation and volumetry of tumors in the liver, lymph nodes and elsewhere will meet or exceed performance in the lung.

Confidence Thresholds:
The 95% confidence thresholds (**±**24%, **±**29%, **±**39%) in Claims 1, 2 and 3 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):
The existence of a true change is described in Claims 1, 2 and 3 in terms of the minimum measured change required to be 95% confident a change has occurred. So, 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).

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. (See Confidence Interval of Result in section 3.1.2 below). If you measured the volume to be 34 cm^3 at baseline and 268 cm^3 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 $cm^3 \pm 45$. A confidence interval that contains zero indicates one should not conclude a true change has occurred.

Whether a change 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.

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 appropriate revisions will be 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, 3 and 4 reflect the likely impact of variations permitted by this Profile. The Profile requires that for a given tumor the same conformant 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 conformant 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 2-1: Minimum Detectable Differences for Tumor Volume Changes (Informative)

	Different Acquisition Device				Same Acquisition Device			
	Different Radiologist		**Same Radiologist**		Different Radiologist		Same Radiologist	
Tumor Diameter	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 repeatability or reproducibility coefficient, depending on the column.
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 can be calculated from the following formula: $1.96 \times \sqrt{(2 \times wCV^2)}$, 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 estimates of the sources of variation 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 3-1: Actors and Required Activities

Actor	Activity	Section
Site	Site Conformance	3.0
Acquisition Device	Product Validation	3.1
Reconstruction Software	Product Validation	3.1
Image Analysis Tool	Product Validation	3.1
Radiologist	Staff Qualification	3.2
	Protocol Design	3.4
	Subject Handling	3.5
	Image QA	3.8
	Image Analysis	3.9
Physicist	Periodic QA	3.3
	Protocol Design	3.4
Technologist	Subject Handling	3.5
	Image Data Acquisition	3.6
	Image Data Reconstruction	3.7

Formal claims of conformance by the organization responsible for an Actor shall be in the form of a published QIBA Conformance Statement. QIBA Conformance Statements for Acquisition Devices, Reconstruction Software and Image Analysis Tools shall describe configuration settings or "Model-specific Parameters" (e.g. protocols) used to achieve conformance.

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 conformance 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 conformance 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:

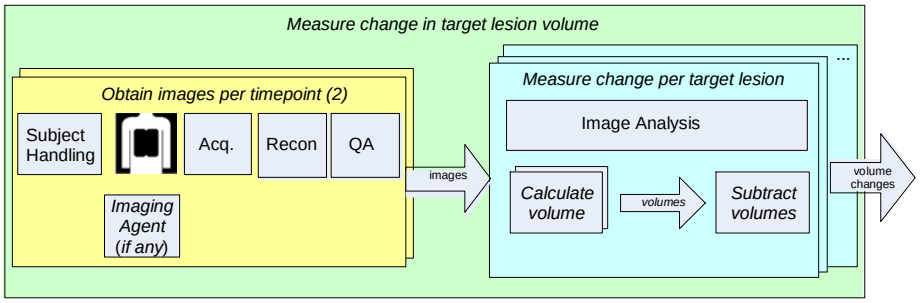

Figure 1: CT Tumor Volumetry – Activity Sequence

The method for measuring change in tumor volume may be described as a pipeline. Subjects are prepared for scanning, raw image data is acquired, and 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 $50cm^3$ and $100cm^3$ 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.

3.0. Site Conformance

This activity involves establishing the overall conformance of an imaging site to this Profile. It includes criteria to confirm the conformance of each of the participating Actors at the site.

3.0.1 DISCUSSION

A site conforms to the Profile if each relevant actor conforms to each requirement assigned ~~to them~~ in the Activities of the Profile. Activities represent steps in the chain of preparing for and generating biomarker values (e.g. product validation, system calibration, patient preparation, image acquisition, image analysis, etc.).

Since a site may assess conformance actor by actor, a checklist document is available which extracts, for convenient reference, all the requirements in this Profile and regroups the requirements by Actor.

Sites may be able to obtain a QIBA Conformance Statement for some actors (e.g. Acquisition Devices) attesting to their conformance to this Profile, rather than the site having to confirm conformance themselves.

3.0.2 SPECIFICATION

Parameter	Actor	Specification
Acquisition Devices	Site	Shall confirm all participating acquisition devices conform to this Profile.

Parameter	Actor	Specification
Reconstruction Software	Site	Shall confirm all participating reconstruction software conforms to this Profile.
Image Analysis Tools	Site	Shall confirm all participating image analysis tools conform to this Profile.
Radiologists	Site	Shall confirm all participating radiologists conform to this Profile.
Physicists	Site	Shall confirm all participating physicists conform to this Profile.
Technologists	Site	Shall confirm all participating technologists conform to this Profile.

3.1. Product Validation

This activity involves evaluating the product Actors (Acquisition Device, Reconstruction Software, and Image Analysis Tool) prior to their use in the Profile (e.g. at the factory). It includes validations and performance assessments that are necessary to reliably meet the Profile Claim.

3.1.1 DISCUSSION

Performance measurements of specific protocols are not addressed here. Those are included in section 3.4.2.

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 Computation is assessed to confirm that the software is computing the volume correctly and confirm there is a reasonable lack of bias at individual timepoints.

Tumor Volume Change Repeatability is assessed to confirm that the software produces sufficiently consistent results over a set of test data. Recall that *repeatability* considers multiple measurements taken under the same conditions (same equipment, parameters, reader, algorithm, etc.) but different subjects, while *reproducibility* considers multiple measurements taken where one or more conditions have changed. So while the Profile Claim is addresses reproducibility, this particular requirement is limited to repeatability.

The target repeatability values were chosen based on the work referenced here:

- Athelogou M, Kim HJ, Dima A, et al., Algorithm Variability in the Estimation of Lung Nodule Volume From Phantom CT Scans: Results of the QIBA 3A Public Challenge. Acad Radiol 2016.
- Buckler AJ, Danagoulian J, Johnson K, et al., Inter-Method Performance Study of Tumor Volumetry Assessment on Computed Tomography Test-Retest Data. Acad Radiol 2015; 22:1–16.
- Fenimore C, Lu ZQ, McNitt-Gray MF, et al., Clinician sizing of synthetic nodules to evaluate CT interscanner effects. RSNA 2012.
- McNitt-Gray MF, Kim GH, Zhao B, et al., Determining the Variability of Lesion Size Measurements from CT Patient Datasets Acquired Under "No Change" Conditions. Transl Oncol 2015 Feb; 8(1):55-64.
- Petrick NP, PhD, Kim HJ, Clunie DA, et al., Comparison of 1D, 2D, and 3D Nodule Sizing Methods by Radiologists for Spherical and Complex Nodules on Thoracic CT Phantom Images. Acad Radiol 2014; 21:30–40.

Methods that calculate volume changes directly without calculating volumes at individual time points are acceptable so long as the results are conformant 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.

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

It is currently unclear whether the provision of a calculator inside the Analysis Tool that takes a wCV value configured by the operator and displays the calculation results alongside measurements constitutes a product claim requiring detailed evidence for the FDA. For this reason the Confidence Interval of Result parameter has been a suggestion and is not a requirement for conformance to the Profile.

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 software version, set-up and configuration parameters used, or to be capable of recording the tumor segmentation boundary as a DICOM Segmentation. Systems based on models should be capable of recording the model and parameters. Currently Analysis Tools are not required to persistently record the volume values and confidence intervals since it is assumed the radiologist will dictate any relevant values into the report.

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

Formatted: Body Text

Formatted: Body Text

3.1.2 SPECIFICATION

Parameter	Actor	Requirement
Acquisition Protocol	Acquisition Device	Shall be capable of ~~storing~~ making validated protocols (designed and validated by the manufacturer and/or by the site) available to the technologist at~~and performing~~ scan times ~~with all the parameters set as specified in section 3.4.2 "Protocol Design Specification"~~.
	Acquisition Device	Shall prepare a protocol conformant with section 3.4.2 "Protocol Design Specification"~~ and validate that protocol as described in section 3.4.2~~.
	Acquisition Device	Shall validate that the protocol achieves an f50 value that is between 0.3 mm^{-1} and 0.5 mm^{-1} for both air and soft tissue edges. See section 4.1. Assessment Procedure: In-plane Spatial Resolution
	Acquisition Device	Shall validate that the protocol achieves: • a standard deviation that is < 60HU. See 4.2. Assessment Procedure: Voxel Noise
~~Image Header~~	~~Acquisition Device~~	~~Shall record in the DICOM image header the actual values for the tags listed in the DICOM Tag column in sections 3.4.2 "Protocol Design Specification".~~
~~Image Header~~	~~Acquisition Device~~	~~Shall record actual timing and triggers in the image header by including the Contrast/Bolus Agent Sequence (0018,0012).~~
~~Image Header~~	~~Acquisition Device~~	~~Shall support recording in the image header (Image Comments (0020,4000) or Patient Comments (0010,4000)) information entered by the Technologist about the acquisition.~~
Reconstruction Protocol	Reconstruction Software	Shall be capable of performing reconstructions and producing images with all the parameters set as specified in section 3.4.2 "Protocol Design Specification".
Image Header	Reconstruction Software	Shall record in the DICOM image header the actual values for the tags listed in the DICOM Tag column in section 3.4.2 "Protocol Design Specification" as well as the model-specific Reconstruction Software parameters utilized to achieve conformance.
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 be able to re-process the first time point (e.g. 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.

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Parameter	Actor	Requirement
		See section 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 or equal to 0.16~~%.~~ • a small subgroup repeatability coefficient of less than 0.21~~%~~ • a large subgroup repeatability coefficient of less than 0.21~~%~~ See section 4.4. Assessment Procedure: Tumor Volume Change Repeatability.
Tumor Volume Bias & Linearity	Image Analysis Tool	Shall be validated to achieve: • an overall tumor volume %bias of less than the Allowable Overall %Bias • a tumor volume %bias for each shape subgroup (spherical, ovoid, lobulated) of less than the Allowable Shape Subgroup %Bias • slope ($\hat{\beta}_1$) between 0.98 and 1.02 The Allowable Overall %Bias and the Allowable Shape Subgroup %Bias are taken from Table 3.1.2-2 based on the overall repeatability coefficient achieved by the Image Analysis Tool using the assessment procedure in section 4.4. See section 4.5 Assessment Procedure: Tumor Volume Bias and Linearity.
Confidence Interval of Result	Image Analysis Tool	~~Shall~~ Is encouraged to calculate and make available to the operator the 95% confidence interval for tumor volume change based on the equation: $(Y_2 - Y_1) \pm 1.96 \times \sqrt{(Y_1 \times wCV_1)^2 + (Y_2 \times wCV_2)^2}$ Where Y_1 and Y_2 is the volume measurement at timepoint 1 and 2, wCV_1 and wCV_2 is the within-nodule coefficient of variation for Y_1 and Y_2 as taken from the following table, D_1 and D_2 is the longest in-plane diameter of the volume at timepoint 1 and 2:
~~Result Recording~~	~~Image Analysis Tool~~	~~Shall record 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.~~

D_1, D_2	10-34mm	35-49mm	50-100mm
wCV_1, wCV_2	0.141	0.103	0.085

Table 3.1.2-2:
Allowable Tumor Volume %Bias based on Repeatability Coefficient

Commented [OK3]: TODO – Pending recalculation by Nancy and Nick

Overall Repeatability Coefficient $\widehat{RC}_p$	Allowable Overall %Bias (RMSE Target: 7.1%)	Allowable Shape Subgroup %Bias (RMSE Target: 7.8%)
0.05%	<6.7%	<7.4%
0.06%	<6.5%	<7.3%
0.07%	<6.3%	<7.1%
0.08%	<6.1%	<6.8%
0.09%	<5.8%	<6.6%
0.10%	<5.5%	<6.3%
0.11%	<5.1%	<5.9%
0.12%	<4.6%	<5.6%
0.13%	<4.1%	<5.1%
0.14%	<3.4%	<4.6%
0.15%	<2.6%	<4.0%
0.16%	<1.1%	<3.2%
0.17%	n/a (failed repeatability)	n/a (failed repeatability)

3.2. Staff Qualification

This activity involves evaluating the human Actors (Radiologist, Physicist, and Technologist) prior to their participation in the Profile. It includes training, qualification or performance assessments that are necessary to reliably meet the Profile Claim.

3.2.1 DISCUSSION

These requirements, as with any QIBA Profile requirements, are focused on achieving the Profile Claim. Evaluating the medical or professional qualifications of participating actors is beyond the scope of this profile.

3.2.2 SPECIFICATION

Note: If the Radiologist has an Image Analyst prepare the measurement contours but the Radiologist still reviews and edits them, then the requirement is to validate the measurement performance of the Radiologist. If the Radiologist completely delegates performing the measurements to an Image Analyst, then requirement is to validate the measurement performance of the Image Analyst.

Formatted: Normal

Parameter	Actor	Specification
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 or equal to 0.16%. • a small subgroup repeatability coefficient of less than 0.21% • a large subgroup repeatability coefficient of less than 0.21% See section 4.4. Assessment Procedure: Tumor Volume Change Repeatability.

3.3. Periodic QA

This activity involves periodic quality assurance of the imaging devices that is not directly associated with a specific subject. It includes calibrations, phantom imaging, performance assessments or validations that are necessary to reliably meet the Profile Claim.

3.3.1 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.

3.3.2 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.

3.4. Protocol Design

This activity involves designing acquisition and reconstruction protocols for use in the Profile. It includes constraints on protocol acquisition and reconstruction parameters that are necessary to reliably meet the Profile Claim.

3.4.1 DISCUSSION

The Profile considers Protocol Design to take place at the imaging site, however, sites may choose to make use of protocols developed elsewhere.

The approach of the specifications here 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 the Conformance Statements for Acquisition Devices and Reconstruction Software 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.

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

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.

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

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.

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 section 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 CT Accreditation Program (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. Since this Profile addresses tumors both in the lung and elsewhere in the torso, the f50 is evaluated for both air and soft tissue edges.

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.

Note that specific constraints are not placed on most of the acquisition and reconstruction parameters in a protocol. It is presumed that significant changes to those parameters would result in non-conformant changes in Noise and Resolution. Changes that do not affect the Noise and Resolution are considered insignificant.

Note also that most modern CT scanners are equipped with Automatic Exposure Control that adjusts the scanner radiation output to achieve pre-determined target noise levels in the images as a function of patient size. The qualification of CT scanner noise needs to account for this provision in that the noise is quantified in a standard size phantom object (such as the CT Accreditation Program phantom from the American College of Radiology) and further as a function of size if there is any concern that the noise performance may be outside compliance for larger sizes.

Commented [OK4]: TODO Review – Text from Ehsan

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 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. Kernel characteristics also interact with acquisition parameters and reconstruction algorithm types; a sharper kernel in a low-dose scan might make a greater difference with an FBP Algorithm than with an Iterative Algorithm. 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, although they may choose to use a protocol provided by the vendor of the acquisition device. The Radiologist is also responsible for ensuring that the protocol has been validated, although the Physicist actor is responsible for performing the validation. 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. Protocol design should be done collaboratively between the physicist and the radiologist with the ultimate responsibility to the radiologist. Some parameters are system dependent and may require special attention from a physicist.

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.	
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 Duration for Thorax	Radiologist	Shall achieve a table speed of at least 4cm per second, if table motion is necessary to cover the required	Table Speed (0018,9309)

Parameter	Actor	Specification	DICOM Tag
		anatomy.	
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.	
Reconstructed Image Thickness	Radiologist	Shall set to between ~~1~~ 0.5mm and 2.5mm (inclusive).	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).	Spacing Between Slices (0018,0088)
In-plane Spatial Resolution	Physicist	Shall validate that the protocol achieves an f50 value that is between 0.3 mm^{-1} and 0.~~7~~5 mm^{-1}~~.~~ for both air and soft tissue edges. See section 4.1. Assessment Procedure: In-plane Spatial Resolution	
Voxel Noise	Physicist	Shall validate that the protocol achieves: • a standard deviation that is < 60HU. See section 4.2. Assessment Procedure: Voxel Noise	

3.5. Subject Handling

This activity involves handling each imaging subject at each time point. It includes subject handling details that are necessary to reliably meet the Profile Claim.

3.5.1 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 (although the tolerances for oral timing are larger than for intravenous).

Note that using no contrast at both timepoints would be considered to be consistent enhancement at the two timepoints.

If oral contrast is used, it is recommended to record the total volume and type of contrast used. If intravenous contrast is used, it is recommended to r~~R~~ecord~~ing~~ the ~~use and~~ type of contrast, actual ~~dose~~ total volume administered, concentration, injection rate, ~~and~~ delay and whether a saline flush was used. Ideally this should be recorded 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.

The sternum should be positioned over the midline of the table. The Table Height and Centering should be adjusted so that the midaxillary line is at the widest part of the gantry.

Positioning the chest (excluding the breasts) in the center of the gantry improves the consistency of relative attenuation values in different regions of the lung, and should reduce scan-to-scan variation in the behavior

of dose modulation algorithms. The subject should be made comfortable, to reduce the potential for motion artifacts and to facilitate compliance with breath holding instructions.

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. An approach that promotes scan-to-scan consistency is essential.

When imaging head and neck tumors, it is not unusual to use gantry tilt, or positioning aids to adjust the slice orientation in the head and neck. Again, it is important to achieve reasonable consistency over timepoints for a given patient.

The Subject Handling specification does not place requirements on patient positioning directly, but rather has the radiologist disqualify measurements from the profile when the positioning at the two time points is different. Consistent positioning will help ensure the majority of studies are conformant and thus achieve the profile Claim.

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

Bismuth breast shields (used by some to reduce radiation exposure in the diagnostic CT setting) increase image noise. The effects of breast shields on image quality may vary depending on the types of shields and their positioning on the chest. The American Association of Physicists in Medicine currently does not endorse the use of breast shields, recommending the use of other dose reduction methods, such as dose modulation techniques, instead (https://www.aapm.org/publicgeneral/BismuthShielding.pdf). Thus, the use of breast shields is not recommended for this Profile. If used, position things such as breast shields so they do not degrade the reconstructed images.

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.

Formatted: Font: Bold

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.

Efforts should be made to obtain consistent, reproducible, maximal inspiratory lung volume on all scans. The use of live breathing instructions given at a pace easily tolerated by the patient is strongly recommended. However, depending on local practice preference and expertise, the use of prerecorded breathing instructions may provide acceptable results. Compliance with breathing instructions should be monitored by carefully observing the movement of the chest wall and abdomen to insure that the breathing cycle stays in phase with the verbal instructions. The scan should not be initiated until maximal inspiratory volume is reached and all movement has ceased.

To promote patient compliance, performing a practice round of the breathing instructions prior to moving the patient into the scanner also is strongly recommended. This will make the subject familiar with the procedure, make the technologist familiar with the subject's breathing rate, and allow the technologist to address any subject difficulties in following the instructions.

Sample breathing instructions:

1. "Take in a deep breath" (watch anterior chest rise)
2. "Breathe all the way out" (watch anterior chest fall)
3. "Now take a deep breath in…..in……in…..in all the way as far as you can"
4. When chest and abdomen stop rising, say "Now hold your breath".
5. Initiate the scan when the chest and abdomen stop moving, allowing for the moment it takes for the diaphragm to relax after the glottis is closed.
6. When scan is completed, say "You can breathe normally"

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.

3.5.2 SPECIFICATION

Parameter	Actor	Specification
Contrast Protocol	Radiologist	Shall prescribe a contrast protocol (which may be No Contrast) 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.~~

Parameter	Actor	Specification
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.).~~
~~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).~~

3.6. Image Data Acquisition

This activity involves the acquisition of image data for a subject at either time point. It includes details of data acquisition that are necessary to reliably meet the Profile Claim.

3.6.1 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.

Acquisition Protocols are often selected by the technologist at scan time based on the procedure requested in the modality worklist. For the measurements to be conformant, this Profile requires that the protocol used has been validated (e.g. by a physicist) to meet certain requirements and performance metrics (see Section 3.4.2). The site will need to find some way to communicate to the technologist which protocols have been validated. This may be something in the protocol names, or a paper list for the technologist to consult, or a special pick-list on the modality console. Or a site may, for example, validate ALL protocols for a given procedure so that any protocol the technologist selects will have been validated.

Formatted: Font: Not Bold

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. See Section 3.4 "Protocol Design".

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.

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. A vertical scan plane (no tilt) is expected for all imaging except some head and neck exams.

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

3.6.2 SPECIFICATION

Parameter	Actor	Specification	DICOM Tag
Acquisition Protocol	Technologist	Shall select a protocol that has been previously prepared and validated for this purpose (See section 3.4.2 "Protocol Design Specification"). ~~Shall report if any parameters are modified beyond the specifications in section 3.4.2 "Protocol Design Specification".~~	
~~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~~ Localizer	Technologist	Shall confirm on the localizer (scout) image ~~scanogram~~ the absence of artifact sources that could affect the planned volume acquisitions or alter the attenuation of lung nodules.	
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)~~
~~Image Header~~	~~Technologist~~	~~Shall enter on the console any factors that adversely influenced subject positioning or limited their ability to cooperate (e.g., breath hold, remaining motionless, agitation in subjects with decreased levels of consciousness, subjects with chronic pain syndromes, etc.).~~	~~Image Comments (0020,4000) or Patient Comments (0010,4000~~
~~Acquisition Field of View (FOV)~~	~~Technologist~~	~~Shall set Consistent with baseline.~~	~~Data Collection Diameter (0018,0090)~~

Formatted Table

3.7. Image Data Reconstruction

This activity involves the reconstruction of image data for a subject at either time point. It includes criteria and procedures related to producing images from the acquired data that are necessary to reliably meet the Profile Claim.

3.7.1 DISCUSSION

Note that the requirement to "select a protocol that has been prepared and validated for this purpose" is not asking the technologist to scan phantoms before every patient, or to validate the protocol themselves. Sites are required in section 3.4.2 to have validated the protocols that the technologist will be using and conformance with the protocol depends on the technologist selecting those protocols.

Reconstruction Protocol affects the image pixel characteristics. The selection and reporting requirements imply a need for a method to record and communicate the protocol selected and any significant modifications and make that information available to the radiologist for the QA Activity. The Profile does not dictate any specific method. Manual methods are acceptable.

, but as long as the protocol has been validated to achieve the required image characteristics then they are considered interchangeable.

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, however, since the kernel is configured in the protocol, it's effect on the spatial resolution will have been evaluated by the f50 requirement in the Protocol Design activity (See 3.4.2).

3.7.2 SPECIFICATION

Parameter	Actor	Specification	DICOM Tag
Reconstruction Protocol	Technologist	Shall select a protocol that has been previously prepared and validated for this purpose (See section 3.4.2 "Protocol Design Specification"). ~~Shall report if any parameters are modified beyond those specifications.~~	
~~In-plane Spatial Resolution~~	~~Technologist~~	~~Shall either~~ ~~• select the same protocol as used for the baseline scan, or~~ ~~• select a protocol with a recorded f50 value within 0.2 mm^{-1} of the f50 value recorded for the baseline scan protocol.~~	

Formatted: Font: Bold

Formatted Table

Formatted: Normal, No bullets or numbering

Parameter	Actor	Specification	DICOM Tag
		~~See section 3.4.2 for further details.~~	
~~Voxel Noise~~	~~Technologist~~	~~Shall either~~ ~~• select the same protocol as used for the baseline scan, or~~ ~~• select a protocol with a recorded standard deviation within 5HU of the standard deviation recorded for the baseline scan protocol.~~ ~~See section 3.4.2 for further details.~~	
Reconstructed Image Thickness	Technologist	Shall set to between ~~1.~~0.5mm and 2.5mm (inclusive) if not set in the protocol~~and consistent (i.e. within 0.5mm) with baseline~~.	
Reconstructed Image Interval	Technologist	Shall set to less than or equal to the Reconstructed Image Thickness (i.e. no gap, may have overlap) and consistent with baseline.	
~~Reconstruction Characteristics~~	~~Technologist~~	~~Shall set the reconstruction 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)~~
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)

Formatted Table

3.8. Image QA

This activity involves evaluating the reconstructed images prior to image analysis. It includes image criteria that are necessary to reliably meet the Profile Claim.

3.8.1 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 conformance 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 positioning variation refers to differences in patient orientation (prone, supine, decubitus, etc.) and the use of positioning wedges. If the patient is supine at one time point and prone at another, then the direction of gravity changes and some tumors may deform differently in a cavity, be compressed differently by other structures, or be affected by deformations of the organ in which they are sited.

Scan Plane variation refers to differences in gantry tilt or differences in head/neck positioning. Since several factors that affect volumetry are not isotropic, changing the orientation of the tumor relative to the scan plane from one time point to another can increase variability.

Formatted: Font: Bold

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 section 3.5, 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.

Contrast Enhancement is required to be consistent between the two timepoints. A non-contrast scan at both timepoints satisfies that requirement.

Formatted: Font: Not Bold

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 timepoints, 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 conform with the specifications in this section, that does not affect other tumors in the same study which do conform with these specifications at both timepoints. Further, if a future study results in the excluded tumor being conformant at two timepoints, then the claim holds across those two timepoints.

While the radiologist is responsible for confirming case conformance with the Image QA specifications in Section 3.8.2, 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").

3.8.2 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 cm^3 and 524 cm^3.)
Tumor Margin Conspicuity	Radiologist	Shall confirm the tumor margins are sufficiently conspicuous and unattached to other structures of equal density to distinguish the

Parameter	Actor	Specification
		volume of the tumor.
Contrast Enhancement	Radiologist	Shall confirm that the phase of enhancement, if any, and degree of enhancement ~~of appropriate reference structures (vascular or tissue)~~ are consistent with baseline.
Patient Positioning Consistency	Radiologist	Shall confirm that any tumor deformation due to patient positioning is consistent with baseline (e.g. tumors may deform differently if the patient is supine in one scan and prone in another).
Breath Hold Consistency	Radiologist	Shall confirm that the breath hold state and degree of inspiration is consistent with baseline.
Scan Plane Consistency	Radiologist	Shall confirm that the anatomical slice orientation (due to gantry tilt or patient head/neck repositioning) is consistent with baseline.
Reconstructed Image Thickness	Radiologist	Shall confirm that the reconstructed image thickness is between 0.5mm and 2.5mm, and consistent with baseline (e.g. within 0.5mm).
Field of View	Radiologist	Shall confirm that the image field of view (FOV) resulting from acquisition and reconstruction settings appears 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-conformant to the Profile.
~~Consistency with Baseline~~	~~Radiologist~~	~~Shall confirm that the tumor is similar in both timepoints in terms of all the above parameters.~~

3.9. Image Analysis

This activity involves measuring the volume change for subjects over one or more timepoints. It includes criteria and procedures related to producing quantitative measurements from the images that are necessary to reliably meet the Profile Claim.

3.9.1 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.

The profile requires that the same Image Analysis Tool and the same Radiologist measure 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 2-1 and the related Discussion for more details.

The Analysis Tool is required (See section 3.1.2) to present to the Radiologist for each volume change the Confidence Interval of Result which indicates 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.

3.9.2 SPECIFICATION

Parameter	Actor	Specification
Reading Paradigm	Radiologist	Shall re-process the first timepoint if it was processed by a different Image Analysis Tool or Radiologist.
Result Verification	Radiologist	Shall review & approve margin contours produced by the tool.

4. Assessment Procedures

To conform to this Profile, participating staff and equipment ("Actors") shall support each activity assigned to them in Table 3-1.

To support an activity, the actor shall conform to the checklist of requirements (indicated by "shall language") listed in the Specification table of that activity 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 references an Assessment Procedure subsection here in Section 4.

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 or the AAPM TG233 phantom, 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 3.4.2, 3.6.2 and 3.7.2). The same protocol and parameters shall be used when performing the assessments in section 4.1 and 4.2, i.e., the noise level during resolution assessment should correspond to that measured during noise assessment.

Commented [OK5]: Per Ehsan

The phantom shall be positioned with the center of the phantom at isocenter and properly aligned along the z-axis. For further details, refer to Section C, Step 3 of the CT Accreditation Testing Instructions: ~~as described in the ACR CTAP documentation about alignment of the beads.~~ http://www.acraccreditation.org/~/media/ACRAccreditation/Documents/CT/CT-Accreditation-Testing-Instructions.pdf

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 assessor shall also generate the MTF curve and determine the f50 value using the edge of the "air insert" (i.e. an empty cutout in the phantom). If the phantom does not have a cutout that provides an air edge to assess, it is permitted to use the edge of the phantom.

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.

Thise assessment~~test~~ procedure ~~described here may be~~ is applicable~~ed~~ to ~~both~~ conventional filtered backprojection ~~reconstruction methods and~~and to iterative reconstruction ~~methods~~.

Note that in addition to the x-y plane MTF, the AAPM TG233 phantom and software also provides an axial resolution measurement (MTF in the z-direction), which may be used as a confirmation of the axial resolution anticipated from the reconstructed image thickness.

Commented [OK6]: Discuss if the methodology can also be "or equivalent".

Commented [OK7]: Proposal (Kirsten) to use an assessment procedure with simpler phantom procedure for FBP. (e.g. the lp/cm)

Commented [OK8]: TODO – clarify what we are expecting them to compare and whether there is a normative requirement.

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 conformant with this Profile

(See Section 3.4.2, 3.6.2 and 3.7.2). The same protocol and parameters shall be used when performing the assessments in section 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.

A~~n approximately circular~~ region of interest (ROI) of at least 400 mm^2 shall be placed near the center of the phantom. The assessor shall record the values reported for the ROI mean and standard deviation.

The assessor is encouraged to record and retain the images and associated measurement details but it is not required beyond the two values listed above. Such details can be helpful when the voxel noise is close to the acceptable limit.

Note that noise is assessed here in a standard sized object. In cases of protocols adaptive to the patient size (such as those using Automatic Exposure Control), the qualification of CT scanner noise should include noise as a function of size (using phantom such as that provisioned in AAPM TG233) if there is any concern that the noise performance may be outside compliance for different sizes.

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.

Thi~~s~~ assessment~~test~~ procedure ~~described here~~ is intended to be a simple phantom measurement that can be used to set a reasonable ~~ceiling~~ limit 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 download ~~obtain~~ the test files ~~in DICOM format~~ from the CT Volumetry Profile Conformance Testing section of the QIDW Data Inventory of the Quantitative Imaging Data Warehouse (QIDW http://qidw.rsna.org/) by selecting the LungMan~~. They can be found by searching for the CT volumetry digital reference object~~ (DRO~~) DICOM image set~~.

Note: To access the QIDW, the assessor will be required to register for a (free) user account.

The test files represent a digital ~~test~~ reference 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 LungMan DRO download package also contains an Excel spreadsheet named StudyInfo.xlsx with the tumor location.

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 section 4.4.1).
- Determine the volume change for designated tumors (see section 4.4.2).
- Calculate statistical metrics of performance (see section 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 download ~~obtain~~ the test files ~~in DICOM format~~ by going to the Quantitative Imaging Data Warehouse (QIDW http://qidw.rsna.org/), selecting Collections, selecting QIDW Data Inventory, selecting~~from the~~ CT Volumetry Profile Conformance Testing, ~~section of the Quantitative Imaging Data Warehouse (QIDW http://qidw.rsna.org/) by~~ selecting ~~the test-retest subset of the~~ RIDER Lung CT Data~~set~~ and choosing the compressed CT Data file.

Note: To access the QIDW, the assessor will be required to register for a (free) user account.

Commented [OK9]: There should be a Profile Conformance Testing link for each profile on the front page and it should be downloadable (read only) without creating an account.

Go to QIDW, select CT Volumetry Profile Performance Testing, and click the link for RIDER Lung CT Data.

The test files represent 31 cases, with two time points per case, each with one target tumor to segment. Each timepoint of each case is represented by a set of DICOM files. The scans have multiple nodules of varying sizes. 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 a .csv file associated with each study in the ~~Dataset~~ download package. The RIDER Lung CT Data download package also contains an Excel spreadsheet named Test-retest-Clinical Study-Summary-Lesion-Position.xlsx that summarizes all the tumor locations and will also help you perform the record keeping and calculations later in this assessment procedure. Note that for some of the cases the two timepoints are in different series in the same study and for some of the cases the two timepoints are in different studies.

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.) 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 shall be reported by the assessor. It is undesirable to test using training data, but until more datasets are available it may be unavoidable.

4.4.2 DETERMINE VOLUME CHANGE

Import the DICOM files into the analysis software. The assessor shall segment each target tumor at each timepoint as described in the Image Analysis Activity (See section 3.9). 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 (e.g. to ensure the volumetric assessment incorporates the whole nodule and excludes any adjacent tissues), 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.

Note: Eleven of the 31 cases in the test files do not meet the Image QA criteria specified by the profile (See 3.8.2). These cases are marked as "excluded" on the Results page of the QIBA spreadsheet and are not included in the calculation of performance metrics. Assessors may skip measuring those cases.

The assessor shall calculate the volume (Y) of each target tumor at time point 1 (denoted Y_{i1}) and at time point 2 (Y_{i2}) where ~~*i*~~*l* denotes the *i*-th target tumor.

The assessor shall calculate the resulting % volume change (d) for each target tumor as $d_i = \ln(Y_{i2}) - \ln(Y_{i1})$.

The downloaded QIBA spreadsheet may be used to record the volume measurements and will perform these calculations and the statistical metrics that follow. Recording the amount of time spent on each case and any comments or concerns is not required for the assessment but is appreciated as feedback to the QIBA Biomarker Committee.

4.4.3 CALCULATE STATISTICAL METRICS OF PERFORMANCE

The assessor shall calculate the within-subject Coefficient of Variation (wCV), where N=~~31~~ 20 and

$$wCV = \sqrt{\sum_{i=1}^{N} d_i^2 /(2 \times N)}$$

The assessor shall estimate the Repeatability Coefficient (RC) as
$\widehat{RC} = 2.77 \times wCV$

~~The assessor shall convert the Repeatability Coefficient (RC) estimate to a percentage as = —1) * 100%.~~

The assessor shall divide the target tumors into a small subgroup (containing the ~~15~~ 14 target tumors with the smallest measured volumes; tagged in the spreadsheet) and a large subgroup (containing the ~~1~~6 tumors with the largest measured volumes; tagged in the spreadsheet). 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 ordinary least squares (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~~ download the test files ~~in DICOM format~~ by going to the Quantitative Imaging Data Warehouse (QIDW http://qidw.rsna.org/), selecting Collections, selecting QIDW Data Inventory, selecting~~from the~~ CT Volumetry Profile Conformance Testing,~~section of the Quantitative Imaging Data Warehouse (QIDW) by~~ selecting QIBA Lung Collection and downloading the contents of all the subfolders ~~the FDA-Lungman-N1-data subset of the RIDER Lung CT Dataset.~~

Note: To access the QIDW, the assessor will be required to register for a (free) user account.

Commented [OK10]: Can we make this a little easier? I tried selecting the top folder and using the drop down to "download checked resources" but it didn't seem to do anything.

The test files represent 3 repeated scans of the FDA Lungman N1 phantom on each of 2 CT scanners. Each timepoint of each case is represented by a set of DICOM files. The phantom contains 7 synthetic tumors, each with a different combination of shape and diameter (see Table 4.5.1-1). The list of 7 target tumors and coordinates are provided in a .csv file associated with each ~~study~~ synthetic tumor in the ~~Dataset~~ download package. The QIBA Lung Collection download package also contains an Excel spreadsheet named StudyInfo.xlsx that summarizes all the tumor locations and will also help you perform the record keeping

and calculations later in this assessment procedure. Note that the images contain half a dozen or so additional tumors that are not identified in the .csv file. Do NOT include measurements of the additional tumors in the results or calculations described in sections 4.5.2 & 4.5.3.

Note: the entire QIBA Lung Collection package can be downloaded as a single zip file: QIBA_Lung_Collection-20170302T222624Z-001.zip

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 measurable (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: Pitch: Collimation: Exposure: Slice Thickness: Increment: Filter: Repeat Scans:	120 1.2 16x1.5 100 mAs 2 mm 1 mm Medium 3
Siemens 64	KVp: Pitch: Collimation: Exposure: Slice Thickness: Increment: Filter: Repeat Scans:	120 1.2 64x0.6 100 mAs 1.5 mm 1.5 mm Medium 3

4.5.2 DETERMINE VOLUME

Import the DICOM files into the analysis software. 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 3.9).

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 (e.g. to ensure the volumetric assessment incorporates the whole nodule and excludes any adjacent tissues), 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.

The downloaded QIBA spreadsheet may be used to record the volume measurements and will perform these calculations. Recording the amount of time spent on each case and any comments or concerns is not required for the assessment but is appreciated as feedback to the QIBA Biomarker Committee.

4.5.3 CALCULATE STATISTICAL METRICS OF PERFORMANCE

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

The assessor shall calculate the individual percentage bias (b_i) of the measurement of each target tumor as $b_i = l\sout{n}\sout{o}g Y_i - l\sout{n}\sout{o}g\ X_i$

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

$$\widehat{D} = \sqrt{\sum_{i=1}^{N} b_i\ /N}$$

The assessor shall convert to a percentage bias estimate as

$$\%\widehat{bias} = \left(\exp\left(\widehat{D}\right) - 1\right) \times 100.$$

The assessor shall fit an ordinary least squares (OLS) regression of the $l\sout{n}\sout{o}g\ Y_i$ on $l\sout{n}\sout{o}g\ X_i$ and shall estimate the slope ($\hat{\beta}_1$).

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 ($l\sout{n}\sout{o}g\ Y_i$ versus $l\sout{n}\sout{o}g\ X_i$) and the OLS regression curve of the volume estimates as part of the assessment record.

4.6. Assessment Procedure: Imaging Site Performance

Note: In this Consensus Stage of the Profile, there is no overall performance requirement on the Site. The future Claim Confirmed Stage of the QIBA Profile development process will include measuring the overall site performance and confirming the performance stated in the Profile Claim is achieved. The procedure in this section is an outline of the process that is expected to be used at that time and will include more details in the future.

This procedure can be used by an imaging site to evaluate the combined performance of all the Actors and Activities at the site.

The assessment procedure has the following steps:

- Validate image acquisition (see section 4.6.1).
- Generate a test image set (see section 4.6.2).
- Assess Tumor Volume Change Variability (see section 4.4.2, 4.4.3 above).

The procedure presumes that the Actors being used by the site meet 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).

4.6.1 ACQUISITION VALIDATION

Review patient handling procedures for conformance with Section 3.5

Establish acquisition protocols and reconstruction settings on the Acquisition Device conformant with 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.

4.6.2 TEST IMAGE SET

Locally acquire a test image set using the protocols established and tested in Section 4.6.1.

The test image set should conform to the characteristics described in Section 4.6.1.

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. Consider 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 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: Made revisions to text to try to achieve an appropriate balance. The details of conformance 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 today 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 conformance 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 is appreciated (as is 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 IRB issues if any.
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 conformant but are allowed to be different) or repeatability (actors must be conformant 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 conformant 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 in the future if 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) Biomarker Committee. The committee is composed of representatives from academia, professional societies, imaging device manufacturers, image analysis software developers, image analysis laboratories, biopharmaceutical industry, government research organizations, 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=Committees.

Profile Editor: Kevin O'Donnell, MASc., Toshiba Medical Research Institute-USA, Inc.

Field Code Changed

QIBA CT Volumetry Biomarker Committee Profile Co-Authors:

Samuel G. Armato III, PhD	The University of Chicago
Rick Avila	Accumetra
Hubert Beaumont, PhD	MEDIAN Technologies
Andrew J. Buckler, MS	Elucid Bioimaging, Inc.
David Clunie, MBBS	PixelMed
Vadivel Devaraju, MS, PhD	The University of Mississippi Medical Center
Paul R. Garrett, MD	Smith Consulting, LLC
Marios A. Gavrielides, PhD	U.S. Food and Drug Administration (FDA)
Ritu Gill, MD, MPH	Brigham and women's hospital
Gregory Goldmacher, MD, PhD, MBA	Merck & Co.
Lubomir Hadjiiski, PhD	University of Michigan
Rudresh Jarecha, MBBS, DMRE, DNB	PAREXEL Informatics
Grace Hyun Kim, PhD	University of California, Los Angeles (UCLA)
Claudia Kirsch, MD	Northwell Health
Qin Li, PhD	U.S. Food and Drug Administration (FDA)
Michael McNitt-Gray, PhD	Dept. of Radiological Sciences, David Geffen School of Medicine at UCLA
James L. Mulshine, MD	Rush University
Reginald F. Munden, MD, DMD, MBA	Houston Methodist Hospital and Research Institute
Nancy A. Obuchowski, PhD	Cleveland Clinic Foundation
Kevin O'Donnell, MASc	Toshiba Medical Research Institute-USA, Inc.
Nicholas Petrick, PhD	U.S. Food and Drug Administration (FDA)
Marthony Robins, PhD	Duke University
Ehsan Samei, PhD	Duke University
Jenifer Siegelman, MD, MPH	Harvard Medical School Brigham and Women's Hospital
Daniel C. Sullivan, MD	Duke University
Ying Tang, PhD	CCS Associates, Inc.

Pierre Tervé, MS	Keosys
Amit Vasanji, PhD	ImageIQ
Binsheng Zhao, D.Sc.	Columbia University

QIBA CT Volumetry Biomarker Committee Profile Contributors:

Maria Athelogou, PhD	DEFINIENS AG
Kristin Borradaile, MS	BioClinica, Inc.
Patricia Cole, PhD, MD	Takeda Pharmaceuticals
Heang-Ping Chan, PhD	University of Michigan
James Conklin, MS, MD	ICON Medical Imaging (retired)
Gary Dorfman, MD	Weill Cornell Medical College
Edward A. Eikman, MD	Moffitt Cancer Center & Research Institute at University of South Florida
Charles Fenimore, PhD	Image-Quality Measurement Consultancy
Robert Ford, MD	Clinical Trials Imaging Consulting, LLC
Matthew Fuld, PhD	Siemens Medical Solutions USA, Inc.
Kavita Garg, MD	University of Colorado
Ronald Gottlieb, MD, MPH	University of Arizona
Dave Gustafson, PhD	Perfint Healthcare
Wendy Hayes, DO	Bristol Myers Squibb
Bruce Hillman, MD	University of Virginia Health System
Philip Judy, PhD	Brigham and Women's Hospital (semi-retired)
Gerhard Kohl, PhD	Siemens AG
Oliver Lehner	Definiens AG
John Lu, PhD	National Institute of Standards & Technology
P. David Mozley, MD	Endocyte Inc.
Daniel Nicholson	Definiens AG
J. Michael O'Neal, MD	BioClinica Inc.
Eric S. Perlman, MD	Perlman Advisory Group, LLC
Aria Pezeshk, PhD	U.S. Food and Drug Administration (FDA)
Anthony Reeves, PhD	Cornell University
Samuel Richard, PhD	Duke University
Yuanxin Rong MD, MPH	PAREXEL International
Berkman Sahiner, PhD	U.S. Food and Drug Administration (FDA)
Ganesh Saiprasad, PhD	Medtronic
Lawrence Schwartz, MD	New York Presbyterian Hospital/Columbia University Medical Center
Eliot Siegel, MD	University of Maryland
Michael D. Silver, PhD	Toshiba Medical Research Institute – USA
Neil Steinmetz, MD, JD	Translational Sciences Corporation
Matthias Thorn, PhD	Siemens AG
David Vining, MD	MD Anderson Cancer Center
David Yankelevitz, MD	Mt. Sinai School of Medicine
Hiro Yoshida, PhD	Harvard - Massachusetts General Hospital

Lifeng Yu, PhD	Mayo Clinic Rochester
Luduan Zhang, PhD	LZ Biomedical

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: 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.

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

Appendix E: Conformance Checklists

QIBA Checklist:

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

INSTRUCTIONS

This Checklist is organized by "Actor" for convenience. If a QIBA Conformance Statement is already available for an actor (e.g. your analysis software), you may choose to provide a copy of that statement rather than confirming each of the requirements in that Actors checklist yourself.

Within an Actor Checklist the requirements are grouped by the corresponding Activity in the QIBA Profile document. If you are unsure about the meaning or intent of a requirement, additional details may be available in the Discussion section of the corresponding Activity in the Profile.

Conforms (Y/N) indicates whether you have performed the requirement and confirmed conformance. When responding **N**, please explain why.

Site Opinion is included during the Technical Confirmation process to allow you to indicate how the requirement relates to your current, preferred practice. When responding **Not Feasible** or **Feasible, will not do** (i.e. not worth it to achieve the Profile Claim), please explain why.

Since several of the requirements mandate the use of specific assessment procedures, those are also included at the end to minimize the need of referring to the Profile document.

Feedback on all aspects of the Profile and associated processes is welcomed.

SITE CHECKLIST

Commented [OK11]: TODO – when updates to the specs are done in Part 3, update these checklists to match.

Site Checked:

Parameter	Conforms (Y/N)	Requirement	Site Opinion
Site Conformance (section 3.0)			
Acquisition Devices		Shall confirm all participating acquisition devices conform to this Profile.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Reconstruction Software		Shall confirm all participating reconstruction software conforms to this Profile.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Image Analysis Tools		Shall confirm all participating image analysis tools conform to this Profile.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Radiologists		Shall confirm all participating radiologists conform to this Profile.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Physicists		Shall confirm all participating physicists conform to this Profile.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Technologists		Shall confirm all participating technologists conform to this Profile.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible

Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt

ACQUISITION DEVICE AND RECONSTRUCTION SOFTWARE CHECKLIST

Acquisition Device(s) Checked - Make/Model/Version:

Parameter	Conforms (Y/N)	Requirement	Site Opinion
Product Validation (section 3.2)			
Acquisition Protocol		Shall be capable of making validated protocols (designed and validated by the manufacturer and/or by the site) available to the technologist at scan time.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
		Shall prepare a protocol conformant with section 3.4.2 "Protocol Design Specification".	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
		Shall validate that the protocol achieves an f50 value that is between 0.3 mm^{-1} and 0.5 mm^{-1}. See section 4.1. Assessment Procedure: In-plane Spatial Resolution	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
		Shall validate that the protocol achieves: • a standard deviation that is < 60HU. See 4.2. Assessment Procedure: Voxel Noise	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Image Header		Shall record in the DICOM image header the actual values for the tags listed in the DICOM Tag column in sections 3.4.2 "Protocol Design Specification".	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Image Header		Shall support recording in the image header (Image Comments (0020,4000) or Patient Comments (0010,4000)) information entered by the Technologist about the acquisition.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Reconstruction Protocol		Shall be capable of performing reconstructions and producing images with all the parameters set as specified in 3.4.2 "Protocol Design Specification".	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Image Header		Shall record in the DICOM image header the actual values for the tags listed in the DICOM Tag column in section 3.4.2 "Protocol Design Specification" as well as the model-specific Reconstruction Software parameters utilized to achieve compliance.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible

IMAGE ANALYSIS TOOL CHECKLIST

Image Analysis Tool(s) Checked - Make/Model/Version:

Parameter	Conforms (Y/N)	Requirement	Site Opinion
Product Validation (section 3.2)			
Multiple Tumors		Shall allow multiple tumors to be measured.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Multiple Tumors		Shall either correlate each measured tumor across time points or support the radiologist to unambiguously correlate them.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Reading Paradigm		Shall be able to present the reader with both timepoints side-by-side for comparison when processing the second timepoint.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Reading Paradigm		Shall be able to re-process the first time point (e.g. if it was processed by a different Image Analysis Tool or Radiologist).	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Tumor Volume Computation		Shall be validated to compute tumor volume with accuracy within 3 % of the true volume. See section 4.3 Assessment Procedure: Tumor Volume Computation.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Tumor Volume Change Repeatability		Shall be validated to achieve tumor volume change repeatability with: • an overall repeatability coefficient of less than or equal to 0.16 • a small subgroup repeatability coefficient of less than 0.21 • a large subgroup repeatability coefficient of less than 0.21 See section 4.4. Assessment Procedure: Tumor Volume Change Repeatability.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Tumor Volume Bias & Linearity		Shall be validated to achieve: • an overall tumor volume %bias of less than the Allowable Overall %Bias • a tumor volume %bias for each shape subgroup (spherical, ovoid, lobulated) of less than the Allowable Shape Subgroup %Bias • slope ($\hat{\beta}_1$) between 0.98 and 1.02 The Allowable Overall %Bias and the Allowable Shape Subgroup %Bias are taken from Table 3.1.2-2 based on the overall repeatability coefficient achieved by the Image Analysis Tool using the assessment procedure in section 4.4. See section 4.5 Assessment Procedure: Tumor Volume Bias & Linearity.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible

Parameter	Conforms (Y/N)	Requirement	Site Opinion
Confidence Interval of Result		Shall calculate and make available to the operator the 95% confidence interval for tumor volume change based on the equation: $(Y_2 - Y_1) \pm 1.96 \times \sqrt{(Y_1 \times wCV_1)^2 + (Y_2 \times wCV_2)^2}$ Where Y_1 and Y_2 is the volume measured at timepoint 1 and 2, wCV_1 and wCV_2 is the within-nodule coefficient of variation for Y_1 and Y_2 as taken from the following table, D_1 and D_2 is the longest in-plane diameter of the volume at timepoint 1 and 2: D_1, D_2 | 10-34mm | 35-49mm | 50-100mm wCV_1, wCV_2 | 0.141 | 0.103 | 0.085	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Result Recording		Shall record percentage volume change relative to baseline for each tumor.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Result Recording		Shall record the confidence interval of result for each change measurement.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Result Recording		Shall record the image analysis tool version.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible

Table 3.1.2-2:
Allowable Tumor Volume %Bias based on Repeatability Coefficient

Overall Repeatability Coefficient $\widehat{RC}$	Allowable Overall %Bias (RMSE Target: 7.1%)	Allowable Shape Subgroup %Bias (RMSE Target: 7.8%)
0.05	<6.7%	<7.4%
0.06	<6.5%	<7.3%
0.07	<6.3%	<7.1%
0.08	<6.1%	<6.8%
0.09	<5.8%	<6.6%
0.10	<5.5%	<6.3%
0.11	<5.1%	<5.9%
0.12	<4.6%	<5.6%
0.13	<4.1%	<5.1%
0.14	<3.4%	<4.6%
0.15	<2.6%	<4.0%
0.16	<1.1%	<3.2%
0.17	n/a (failed repeatability)	n/a (failed repeatability)

RADIOLOGIST CHECKLIST

Note: The Radiologist is responsible for the protocol parameters, although they may choose to use a protocol provided by the vendor of the acquisition device. The Radiologist is also responsible for ensuring that the protocol has been validated, although the Physicist actor is responsible for performing the validation. Protocol design should be done collaboratively between the physicist and the radiologist with the ultimate responsibility to the radiologist. Some parameters are system dependent and may require special attention from a physicist.

Radiologist(s) Checked:

Parameter	Conforms (Y/N)	Specification		Site Opinion
Staff Qualification (section 3.1)				
Tumor Volume Change Repeatability		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 or equal to 0.16 • a small subgroup repeatability coefficient of less than 0.21 • a large subgroup repeatability coefficient of less than 0.21 See 4.4. Assessment Procedure: Tumor Volume Change Repeatability.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Protocol Design (section 3.6.2)				
Acquisition Protocol		Shall prepare a protocol to meet the specifications in section 3.4-protocol design.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Acquisition Protocol		Shall ensure technologists have been trained on the requirements of this profile.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Total Collimation Width		Shall set to Greater than or equal to 16mm.	Total Collimation Width (0018,9307)	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
IEC Pitch		Shall set to Less than 1.5.	Spiral Pitch Factor (0018,9311)	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Nominal Tomographic Section Thickness (T)		Shall set to Less than or equal to 1.5mm.	Single Collimation Width (0018,9306)	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Scan Duration for Thorax		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)	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible

Parameter	Conforms (Y/N)	Specification		Site Opinion
Reconstruction Protocol		Shall prepare a protocol to meet the specifications in this table.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Reconstruction Protocol		Shall ensure technologists have been trained on the requirements of this profile.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Reconstructed Image Thickness		Shall set to between 0.5mm and 2.5mm (inclusive).	Slice Thickness (0018,0050)	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Reconstructed Image Interval		Shall set to less than or equal to the Reconstructed Image Thickness (i.e. no gap, may have overlap).	Spacing Between Slices (0018,0088)	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Subject Handling (section 3.8)				
Contrast Protocol		Shall prescribe a contrast protocol (which may be No Contrast) that achieves enhancement consistent with baseline.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Use of intravenous contrast		Shall determine whether the selected contrast protocol, if any, will achieve sufficient tumor conspicuity.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Use of oral contrast		Shall determine whether the selected contrast protocol, if any, will achieve sufficient tumor conspicuity.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Image QA (section 3.8)				
Patient Motion Artifacts		Shall confirm the images containing the tumor are free from artifact due to patient motion.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Dense Object Artifacts		Shall confirm the images containing the tumor are free from artifact due to dense objects, materials or anatomic positioning.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Clinical Conditions		Shall confirm that there are no clinical conditions affecting the measurability of the tumor.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Tumor Size		Shall confirm (now or during measurement) that tumor longest in-plane diameter is between 10 mm and 100 mm.		□ Routinely do already □ Feasible, will do

Parameter	Conforms (Y/N)	Specification	Site Opinion
		(For a spherical tumor this would roughly correspond to a volume between 0.5 cm^3 and 524 cm^3.)	□ Feasible, will not do □ Not feasible
Tumor Margin Conspicuity		Shall confirm the tumor margins are sufficiently conspicuous and unattached to other structures of equal density to distinguish the volume of the tumor.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Contrast Enhancement		Shall confirm that the phase of enhancement, if any, and degree of enhancement of appropriate reference structures (vascular or tissue) are consistent with baseline.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Tumor Measurability		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-conformant to the Profile.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Consistency with Baseline		Shall confirm that the tumor is similar in both timepoints in terms of all the above parameters.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Image Analysis (section 3.9)			
Reading Paradigm		Shall re-process the first time point if it was processed by a different Image Analysis Tool or Radiologist.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Result Verification		Shall review & approve margin contours produced by the tool.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible

PHYSICIST CHECKLIST

Note: 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.

Physicist(s) Checked:

Parameter	Conforms (Y/N)	Requirement	Site Opinion
Periodic QA (section 3.5)			
QC		Shall perform relevant quality control procedures as recommended by the manufacturer. Shall record the date/time of QC procedures for auditing.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Protocol Design (section 3.6.2)			
In-plane Spatial Resolution		Shall validate that the protocol achieves an f50 value that is between 0.3 mm^{-1} and 0.5 mm^{-1}. See section 4.1. Assessment Procedure: In-plane Spatial Resolution	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Voxel Noise		Shall validate that the protocol achieves: • a standard deviation that is < 60HU. See section 4.2. Assessment Procedure: Voxel Noise	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible

TECHNOLOGIST CHECKLIST

Technologist(s) Checked:

Formatted: Left, None, Space Before: 0 pt, After: 0 pt, Don't keep with next

Parameter	Conforms (Y/N)	Specification	Site Opinion
Subject Handling (section 3.8)			
Use of intravenous contrast		Shall use the prescribed intravenous contrast parameters.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Use of intravenous contrast		Shall document the total volume of contrast administered, the concentration, the injection rate, and whether a saline flush was used.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Use of oral contrast		Shall use the prescribed oral contrast parameters.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Use of oral contrast		Shall document the total volume of contrast administered and the type of contrast.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Subject Positioning		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 in section 3.5.1.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Artifact Sources		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.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Table Height & Centering		Shall adjust the table height for the mid-axillary plane to pass through the isocenter.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Table Height & Centering		Shall position the patient such that the “sagittal laser line” lies along the sternum (e.g. from the suprasternal notch to the xiphoid process).	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Breath hold		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.	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Breath hold		Shall ensure that for each tumor the breath hold state is consistent with baseline.	□ Routinely do already □ Feasible, will do □ Feasible, will not do

Parameter	Conforms (Y/N)	Specification		Site Opinion
				□ Not feasible
Contrast-based Acquisition Timing		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).		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Contrast-based Acquisition Timing		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).		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Image Data Acquisition (section 3.6)				
Acquisition Protocol		Shall select a protocol that has been previously prepared and validated for this purpose (See section 3.4.2 "Protocol Design Specification").		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Acquisition Protocol		Shall report if any parameters are modified beyond the specifications in section 3.4.2 "Protocol Design Specification".		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Scan Plane (Image Orientation)		Shall set Consistent with baseline.	Gantry/Detector Tilt (0018,1120)	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Tube Potential (kVp)		Shall set Consistent with baseline (i.e. the same kVp setting if available, otherwise as similar as possible).	KVP (0018,0060)	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Scanogram		Shall confirm on the scanogram the absence of artifact sources that could affect the planned volume acquisitions.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Scan Duration for Thorax		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)	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Image Header		Shall enter on the console any factors that adversely influenced subject positioning or limited their ability to cooperate (e.g., breath hold, remaining motionless, agitation in subjects with decreased levels of consciousness, subjects with chronic pain syndromes, etc.).	Image Comments (0020,4000) or Patient Comments (0010,4000	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Acquisition Field of View (FOV)		Shall set Consistent with baseline.	Data Collection Diameter (0018,0090)	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible

Parameter	Conforms (Y/N)	Specification		Site Opinion
Image Data Reconstruction (section 3.7)				
Reconstruction Protocol		Shall select a protocol that has been previously prepared and validated for this purpose (See section 3.4.2 "Protocol Design Specification"). Shall report if any parameters are modified beyond those specifications.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Reconstructed Image Thickness		Shall set to between 0.5mm and 2.5mm (inclusive) if not set in the protocol.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Reconstructed Image Interval		Shall set to less than or equal to the Reconstructed Image Thickness (i.e. no gap, may have overlap) and consistent with baseline.		□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Reconstruction Characteristics		Shall set the reconstruction 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)	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible
Reconstruction Field of View		Shall ensure the Field of View spans at least the full extent of the thoracic and abdominal cavity, but not substantially greater than that.	Reconstruction Field of View (0018,9317)	□ Routinely do already □ Feasible, will do □ Feasible, will not do □ Not feasible