QIBA profile conformance testing
CT Tumor Volume Change

Supplement

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 CT_Vol_Checklist document.

Although most of the requirements described in Checklist 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%. Thus, 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 3.5.2, 3.7.2 and 3.8.2). 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.

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: 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 or equivalent 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 internal air edge to assess, it is permitted to use the outer edge of the phantom.

The procedure described above is provided as a reference method. This reference method and the method used by the particular 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.

This assessment procedure is applicable to conventional filtered back-projection and to iterative reconstruction.

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).

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 includes 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 3.5.2, 3.7.2 and 3.8.2). 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, likely close to the center, from the uniformity portion of the phantom.

A region of interest (ROI) of at least 400 mm^2 shall be placed near the center of the slice. 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.

This assessment procedure is intended to be a simple phantom measurement that can be used to set a reasonable 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 Accuracy

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 accurately. 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 the test files by going to the Quantitative Imaging Data Warehouse (QIDW http://qidw.rsna.org/), selecting QIDW Data Inventory, selecting CT Volumetry Profile Conformance Testing, and downloading the LungMan DRO zip file.

> Note: The assessor will not be permitted to access the QIDW Data Inventory until they have registered for a (free) user account and logged in.

The test files include 11 DICOM sequential images representing a digital reference object (a "virtual phantom") with z-axis resolution of 1.5mm. A spherical "tumor" and a box-shaped "tumor", both with -10 HU radio-density, are 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 tumor is segmented.

The assessor shall use the Image Analysis Tool to segment both the spherical tumor and the box-shaped tumor present in the test images and calculate the volume of each tumor.

The assessor shall record the percentage difference between the reported volume and the true value.

The downloaded zip file contains an Excel spreadsheet named "QIBA Volumetry CT - 4.3 Assessment Procedure Tumor Volume Computation" with the coordinates of the centroid of each tumor, the true value for its volume, and statistical analysis tooling to record the results and assess the performance.

4.4. Assessment Procedure: Tumor Volume 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 at two time points. 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 for designated tumors at two time points (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 the test files by going to the Quantitative Imaging Data Warehouse (QIDW http://qidw.rsna.org/), selecting QIDW Data Inventory, selecting CT Volumetry Profile Conformance Testing, selecting RIDER Lung CT Data, and downloading the RIDER Lung CT Data zip file (roughly 4GB).

> Note: The assessor will not be permitted to access the QIDW Data Inventory until they have registered for a (free) user account and logged in.

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. The list of target tumors and coordinates are provided in a .csv file associated with each study in the download package. The RIDER Lung CT Data download package also contains an Excel spreadsheet named "QIBA CTVol TumorVolumeChange Assessment4.4-Repeatability" that summarizes all the tumor locations and will also help the assessor 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 measurable (as defined in the Profile) and have a range of volumes, shapes and types to be representative of the scope of this 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

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 3.10). 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 3.9.2). 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* 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=20 and

$$wCV = \sqrt{\sum_{i=1}^{N} d_i^2 \,/N}$$

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

The assessor shall divide the target tumors into a small subgroup (containing the 14 target tumors with the smallest measured volumes; tagged in the spreadsheet) and a large subgroup (containing the 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.

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 download the test files by going to the Quantitative Imaging Data Warehouse (QIDW http://qidw.rsna.org/), selecting QIDW Data Inventory, selecting CT Volumetry Profile Conformance Testing, and downloading the QIBA Lung Collection zip file (roughly 1GB).

Note: The assessor will not be permitted to access the QIDW Data Inventory until they have registered for a (free) user account and logged in.

The test image set consists of scans of the FDA Lungman N1 phantom using two different scanners from different vendors. Several phantom configurations, using a set of 7 synthetic tumors, each with a different combination of size, shape and diameter (see Table 4.5.1-1), were scanned. The scan of a configuration is repeated 3 times, each resulting in a set of DICOM files. The list of target tumors and centroid coordinates for each scan are provided in an Excel spreadsheet named "QIBA Volumetry CT - 4.5 Tumor volume bias and linearity" in the QIBA Lung Collection download package. The spreadsheet also helps the assessor perform the record keeping and calculations later in this assessment procedure.

Note that the images contain additional tumors that are not identified in the .csv files. Do NOT include measurements of the additional tumors in the results or calculations described in sections 4.5.2 & 4.5.3.

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:	120 1.2 16x1.5 100 mAs 2 mm 1 mm

	Filter:	Medium
	Repeat Scans:	3
Siemens 64	KVp:	120
	Pitch:	1.2
	Collimation:	64x0.6
	Exposure:	100 mAs
	Slice Thickness:	1.5 mm
	Increment:	1.5 mm
	Filter:	Medium
	Repeat Scans:	3

248

249 4.5.2 DETERMINE VOLUME

250 For each scan, the assessor shall import the DICOM files into their analysis software and segment the
251 tumors identified in the spreadsheet as described in the Image Analysis Activity (See 3.10). In total, the
252 assessor will do 39 target tumor segmentations (3 scans each for 7 tumors on 1 scanner and 6 tumors on
253 the other scanner)
254
255 The assessor is permitted to edit the tumor segmentation or seed point if that is part of the normal
256 operation of the tool. If segmentation edits are performed (e.g. to ensure the volumetric assessment
257 incorporates the whole nodule and excludes any adjacent tissues), results shall be reported both with and
258 without editing.
259
260 When evaluating an Image Analysis Tool, a single reader shall be used for this entire assessment procedure.
261 When evaluating a Radiologist, a single tool shall be used for this entire assessment procedure.
262
263 The assessor shall calculate the volume (Y) of each target tumor (denoted Y_i) where *i* denotes the *i*-th target
264 tumor.
265
266 The downloaded QIBA spreadsheet may be used to record the volume measurements and will perform
267 these calculations. Recording the amount of time spent on each case and any comments or concerns is not
268 required for the assessment but is appreciated as feedback to the QIBA Biomarker Committee.

269 4.5.3 CALCULATE STATISTICAL METRICS OF PERFORMANCE

270 The natural log of the true volumes (X_i) of each target tumor are known and are provided in the dataset.
271
272 The assessor shall calculate the individual percentage bias (b_i) of the measurement of each target tumor as
273 $b_i = lnY_i - ln\ X_i$
274
275 The assessor shall estimate the population bias over the N target tumors as
276 $\widehat{D} = \sum_{i=1}^{N} b_i\ /N$
277
278 The assessor shall convert to a percentage bias estimate as
279 $\%\widehat{bias} = \left(\exp\left(\widehat{D}\right) - 1\right) \times 100.$
280
281 The assessor shall estimate 95% confidence intervals for the population bias as

$CI_{\hat{D}} = \hat{D} \pm t_{[\alpha=0.025, df=N-1]} \times SE(\hat{D})$

and

$CI_{\%bias} = (\exp(CI_{\hat{D}}) - 1) \times 100$

The Overall $\%\widehat{bias}$ metric used to compare to the bias specifications is

$Overall\ \%\widehat{bias} = \max(|CI_{\%bias}|)$

The assessor shall fit a quadratic model to the volume data $\log Y_i$ on $[\ln X_i, (\ln X_i)^2]$ and shall estimate the quadratic term ($\hat{\beta}_2$).

The assessor shall fit an ordinary least squares (OLS) regression of the $\ln Y_i$ on $\ln X_i$ 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 ($\ln Y_i$ versus $\ln X_i$) and the OLS regression curve of the volume estimates as part of the assessment record.