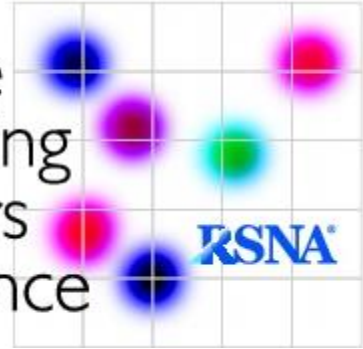


Quantitative
Imaging
Biomarkers
Alliance



QIBA Profile:
Computed Tomography: Lung
Densitometry

5

Table of Contents

	Change Log:.....	4
10	Open Issues:.....	6
	Closed Issues:.....	8
	1. Executive Summary.....	10
	2. Clinical Context and Claims.....	11
	3. Profile Activities.....	13
15	3.1. Scanner Qualification.....	15
	3.1.1 Discussion.....	15
	3.1.2 Specification.....	16
	3.2. Staff Qualification	18
	3.2.1 DISCUSSION.....	18
20	3.2.2 SPECIFICATION.....	18
	3.3. Periodic QA.....	18
	3.3.1 Discussion.....	18
	3.3.2 Specification.....	19
	3.4. Protocol Design	19
25	3.4.1 DISCUSSION.....	19
	3.4.2 SPECIFICATION.....	21
	3.5. Subject Handling.....	23
	3.5.1 Discussion.....	23
	3.5.2 Specification.....	24
30	3.6. Image Data Acquisition.....	25
	3.6.1 Discussion.....	25
	3.6.2 Specification.....	25
	3.7. Image Data Reconstruction.....	26
	3.7.1 Discussion.....	26
35	3.7.2 Specification.....	27
	3.8. Image QA.....	27
	3.8.1 Discussion.....	27
	3.8.2 Specification.....	27
	3.9. Image Distribution.....	28
40	3.10. Image Analysis.....	28
	3.10.1 Discussion.....	28
	3.10.2 Specification.....	29
	3.11. Image Interpretation.....	30
	3.11.1 Discussion.....	30
45	3.11.2 Specification.....	30
	4. Assessment Procedures.....	31
	4.1.1 Assessment Procedure: HU Bias and Repeatability.....	32
	4.1.2 Assessment Procedure: Voxel Noise and Noise Power Spectrum.....	33
	4.1.3 Assessment Procedure: In-Plane, Through-Plane (Z-axis) Spatial Resolution and Edge Enhancement.....	33
50	4.2. Assessment Procedure: Reproducibility of Image Analysis Software across Various	

	Vendors	34
	5. Conformance.....	35
	References	35
55	Appendices.....	38
	Appendix A: Acknowledgements and Attributions.....	38
	Appendix B: Background Information.....	38
	Appendix C: Conventions and Definitions	51
	Appendix D: Model-specific Instructions and Parameters	53
60	Appendix E: Image Analysis Software Comparison	54
	Appendix F. Assessment procedure for computing reproducibility of Lung Density Software across various vendors.....	57
	References for Appendices A-F:.....	59
65		

Change Log:

This table is a best-effort of the authors to summarize significant changes to the Profile.

70

Date	Sections Affected	Summary of Change
2016.06.21	All	First Draft
2016.07.20	Up to Section 3	Per call discussion, modified specification for lung inflation volume. Completed changes through Section 2 with draft to be reviewed for internal comment.
2016.07.24	Section 3	Assigned for discussion at next call (tentatively 2016.08.03)
2016.08.31	All	Cleanup of current discussions, additions to Section 3.
2016.09.26	Discussed Thru Section 3.2	Took comments during call on organization, phantom specifications and statistics.
2016.11.16	Updated Section 4.0, Assessment Procedures	Inserted text documenting detailed harmonization calculations provided by Dr. Chen-Mayer
2016.11.16	Addressed Call Comments Through Section 3	Final approval from group on call.
2017.02.01	Updates to Calibration Sections for Scanner Qualification (3.1)	Stephen Humphries
2017.03.29	Revisions to Clinical Interpretation Sections and to finalize profile	David Lynch and Sean Fain
2017.08.13	<ol style="list-style-type: none"> 1. Added example of clinical context of claims 2. AEC moved to closed issue. 3. 3.3.2 and 3.6.2 specifications made consistent with text 4. References 	Sean Fain - Modifications in response to face to face meeting suggestions and 6/28/17 call.

	<p>updated.</p> <p>5. Appendix D added</p> <p>6. Modifications to section 3.8 (software analysis)</p> <p>7. Created Section 4.3 and place holder for Appendix E.</p>	
2017.08.30	<p>1. Moved AEC back to "Open Issue"</p> <p>2. Created a Clean Version for Distribution</p>	
2017.11.19	Addressed concerns from external/internal review (Kevin O'Donnell, Nancy Obuchowski, and Phil Judy).	
2018.03.04	Final Edits from Committee Discussions	S. Fain and Jered Sieren: Draft for internal distribution to committee in anticipation of distribution for public comment
2018.07.30	Additional Edits incorporated from committee discussion per Dr. Lynch	Modified "shall's" to recommended best methods. Changed radiologist actor to technologist for routine specifications.
2018.11.26	Revised draft distributed for RSNA discussion	Mainly clean up of edits to guide discussion and incorporation of software comparison guidelines.
2018.12.15	Revised draft SBF	Incorporated edits and software comparison updates from Miranda Kirby.
2019.02.25	Revised draft SBF	Incorporated suggestions from committee to be more specific about the use of the COPDGene phantom throughout qualification sections. Addressed potential bias in HU near -1000 due to truncation by adding footnotes to claims and specification tables where appropriate. Fixed typos in sections 4.1.1.; Updated

		section 4.2 to more clearly describe procedure for qualifying a new vendor software analysis tool.
2019.04.23	Revised draft SBF	Incorporating final changes to Appendices E and F and modified qualification of analysis software in Table 3.10 and Section 4.2.
2019.04.25	Revised draft SBF	Incorporating edits software comparison from 4/24/19 call: 1) require deterministic analysis for times 1 and 2; 2) retain appendix F as a reference tool for new analysis software to compare against existing software on a common reference data set. 3) Update all specification tables (missed 3.1.2). 4) indexing edits to Appendices E and F. 5) Fix referencing to have a complete listing of references separately for main and appendices sections of the document.
2019.06.13	SBF	Edits incorporated prior to committee vote

Open Issues:

75 The following issues are provided here to capture associated discussion, to focus the attention of reviewers on topics needing feedback, and to track them so they are ultimately resolved. In particular, comments on these issues are highly encouraged during the Public Comment stage.

<p>1. “Iterative Reconstruction”</p> <p>Q. What is the effect of iterative reconstruction (IR) methods on RA-950 HU and Perc15 measures of lung parenchymal density?</p> <p>A. The resolution of this question is actively under investigation as part of a development/ground work project. The use of IR would be desirable to reduce CT dose to research subjects and patients undergoing quantitative CT of the lungs. Several published works have emerged, one in Medical Physics from this committee, demonstrating that IR methods reduce noise and have non-linear effects on texture and other low contrast structures. Further complicating this issue, vendors use different statistical and model-based IR methods in their commercial software that may affect image noise differently and would thus need to be harmonized across vendors. Additionally, IR methods are likely to continue to evolve, and continued ongoing assessment will be needed. We anticipate that after further study, recommendations for integration of IR methods into the Profile can be added to later versions.</p>
<p>2. “Harmonization”</p> <p>Q. What is the best reference standard for harmonizing systematic differences in quantitative CT number (Hounsfield Unit – HU) across scanner make and model?</p>

A. Two rounds of scans conducted with engineers from 4 major vendors (Siemens, GE, Toshiba, and Philips) using the QIBA-SRM phantom to establish bias and precision of HU measure have been completed. A method to empirically correct to a common reference scanner has been shown to reduce bias and improve precision of qCT measures in this phantom. A model-based correction method based on the composition (best-knowledge) and the certified physical densities of the constituent materials of the QIBA-SRM phantom has also shown promise as an absolute correction (standardization) method, and a manuscript is under review by Medical Physics. The standardization method is being actively pursued and tested in the Round 2 analysis, with data acquisition completed and analysis pending. The harmonization method is provisionally included in the current profile as a recommended assessment procedure (Section 4.2) with the goal of addressing a cross-sectional claim in future versions of the profile.

3. “Airway measures”

Q. Does this profile meet any standards for airway morphology (e.g. wall thickness and lumen area) measurement?

A. This profile meets claims for parenchymal density analysis only, but spatial resolution specifications under acquisition and reconstruction are also consistent with current methods in the literature for measurement of airway morphology assuming commercially available software analysis methods. A claim for airway measures is left for future versions of the Profile.

4. “Regional measures”

Q. Should the measurement methods specify regional measures (e.g. upper, middle, lower thirds) or lobar segmentation and density measurement methods used to guide lung volume reduction interventions?

A. There is currently inadequate evidence to evaluate this issue, but there is little doubt that lobar segmentation is an emerging application for targeted lung volume reduction (by valve placement or surgically). While lobar specific density analysis is not part of the current profile, we expect to address this directly in future versions. However, it should be noted that there is no reason to think that the proposed protocols, including acquisition and reconstruction would not still serve this application equally well with the same quality assurance processes, not including software analysis. The main challenge preventing us from including this application in the claims is that some software analysis and measurement steps would need to be performed at the lobar level, raising new challenges for accuracy and precision of segmentation, especially using automated techniques. More effort to characterize consistency and consequences of errors in lobar segmentation for lung volume reduction applications are needed to define claims for this application.

5. “Automated Exposure Control (AEC)”

Q. What is the effect of automated exposure control (AEC) on RA-950 HU and Perc15

across scanner make and model?

A. The use of AEC is desirable as a method that reduces dose and makes noise behavior more consistent throughout the image by matching tube current to achieve similar photon counts across varying structural attenuation. However, different vendors match performance to the selected AEC parameter to emphasize different features in the image and proprietary models are used to predict tube current modulation based on initial scout scans. Ground work performed by the Lung Density Biomarker Committee has led to harmonized protocols that match CT dose for AEC parameter selection across scanner makes and models (see protocol examples in Appendix D). More study is needed to resolve this issue in general, but the committee considers that AEC is sufficiently mature and resolved by empirically matching settings across scanner makes and models for an average sized patient (i.e. 75 kg) for inclusion in study protocols. In the present version, the amount of radiation higher or lower than a target of 3 mGy is based on patient size and shape according to each manufacturer’s AEC attenuation model. CT radiation dose in the chest is expected to vary by $\pm 18\%$ for subject weight (see Huda et al., Med Phys. 2010 Feb;37(2):842-7). between 50-100 kg, which is considered sufficiently small to be within the range of equivalent performance for the current claims as stated. Recommendations will be updated in later versions of this profile to recommend general solutions to support cross-sectional claims.

Closed Issues:

80 The following issues have been considered closed by the biomarker committee. They are provided here to forestall discussion of issues that have already been raised and resolved, and to provide a record of the rationale behind the resolution.

1. “Elastic Clause”

Q. Is this template open to further revisions?

A. Yes.

This is an iterative process by nature.

Submit issues and new suggestions/ideas to the QIBA Process Cmte.

2. “Biomarkers”

Q. Choice of key biomarkers of lung density resolved?

A. Yes, RA-950 HU and Perc15 are the most established measures of emphysema, as both are validated against tissue histology. Perc15, or Perc15 adjusted for lung volume, is used most ubiquitously in clinical research trials.

3. “Regional measures”

Q. Whole lung or regional (e.g. lobar) measures of density recommended?

A. Whole lung at present. Regional measures may be introduced in later versions.

4. “Airways measures”

<p>Q. Will airway measures be included?</p> <p>A. No.</p>
<p>5. “Air trapping measures”</p> <p>Q. Will density measures of air trapping be included?</p> <p>A. No.</p>
<p>6. “Lung Volume Adjustment”</p> <p>Q. Will lung volume adjustment of RA-950 HU and Perc15 be included in the Profile?</p> <p>A. Yes, but no specific method will be recommended in this version of the Profile as this remains a developing area of research.</p>
<p>7. “Breath-hold Consistency”</p> <p>Q. There is a concern that a subject with 2 weak efforts, but less than 10% difference in lung inflation would still be accepted by this profile. Is a subject with this type of effort adequately quantifiable by this profile?</p> <p>A. Yes. Based on review of Dr. Park’s study in the meta-analysis, the range of fractional change of volume, V_2/V_1, is [0.90, 1.11], i.e. [-10%, +11%]. This was the study with the longest time interval between baseline and follow-up scans. This case is well addressed by the current Profile in “Section 3.3 Subject Handling” where it is explicitly described that the breath-hold coaching required conforms to a lung inflation standard that would meet the claims. Because we are assessing longitudinal change, it is less important (but still desirable) that the subject be within 90% of vital capacity. Published works do not in practice require spirometric gating (see Gierada et al., Radiology 220(2):448-454.), nor was spirometric gating of breath-hold used in the studies included in the meta-analysis.</p>
<p>8. “Cross-Sectional Claim”</p> <p>Q. Will there be a cross-sectional claim?</p> <p>A. No, there is not sufficient data at the present time to support a cross-sectional claim. Both current claims are longitudinal, reporting change in emphysema extent. Current groundwork testing a harmonization method across scanner make and model, if successful, will make a cross-sectional claim feasible in future versions of the profile.</p>
<p>9. “Specification of matrix size”</p> <p>Q. Should the acquisition parameters include specification of matrix size or display FOV?</p> <p>A. This is specified if the in-plane spatial resolution is met.</p>

1. Executive Summary

90 The goal of a QIBA Profile is to achieve a repeatable and useful level of performance for measures of lung density from quantitative CT using the RA-950 HU and Perc15 biomarkers of emphysema. Please see Appendix B for more detailed information on the calculation of and rationale for RA-950 HU and Perc15 as the biomarkers of choice.

The **Claim** (Section 2) describes the performance in terms of bias and precision of RA-950 HU and Perc15 for detecting change in lung density.

95 The **Activities** (Section 3) describe how to generate RA-950 HU and Perc15 for longitudinal studies of the change in lung density. Requirements are placed on the **Actors** that participate in those activities as necessary to achieve the Claim in Section 2.

Assessment Procedures (Section 4) for evaluating specific requirements are defined as needed.

100 This QIBA Profile (CT: Lung Densitometry) addresses RA-950 HU and Perc15 for longitudinal studies which are often used as biomarkers of emphysema progression in chronic obstructive pulmonary disease (COPD) or as a response to cessation of smoking and possible future treatment approaches. It places requirements on Acquisition Devices, Physicists, Technologists, Clinicians, Statisticians, Reconstruction Software and Image Analysis Software involved in Scanner Qualification, Staff Qualification, Periodic Quality Assurance, Subject Handling, Protocol Design, Image Data Acquisition, Image Data Reconstruction, Image QA, Image Distribution and
105 Image Analysis.

The requirements are focused on achieving negligible bias and avoiding unnecessary variability of the RA-950 HU and Perc15 measurements by compensating for variations in CT number due to inconsistency of lung inflation volume and calibration of the CT scanner, and vendor-specific bias due to CT scanner make and model. To meet the claims, scanner calibration is performed
110 using a well characterized imaging phantom ideally containing lung equivalent density foams as described in Section 4.1.

The clinical performance targets are to achieve bias and repeatability such that a change in RA-950 HU of $\geq 3.7\%$ of the normalized lung volume, or a change in Perc15 of ≥ 11 HU can be accepted as indicative of a true change (with 95% confidence).

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

Note that this document only states requirements to achieve the claim, not “requirements for standard of care.” Conformance to this Profile is less important than providing appropriate
120 patient care.

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

2. Clinical Context and Claims

125 Clinical Context

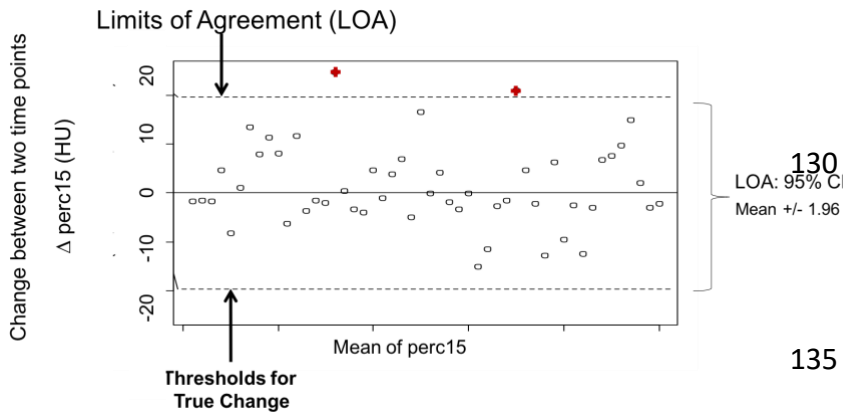


Figure 1: Example Bland-Altman plot for perc15 without VA; the LOA (or RC when there's no bias) are the thresholds required for detection of a true increase in the extent of emphysema. The first step is to assess whether a measurement meets the threshold for a true change (red crosses above the upper LOA would be considered subjects showing true change). The magnitude of the change can be expressed as a 95% CI: $\Delta \pm 18HU$.

The clinical context for this profile includes studies of quantitative longitudinal change in lung parenchymal density using image processing of CT scans acquired at different time points to quantify progression of emphysema in COPD. These studies specifically evaluate increase or decrease of lung relative area falling below a threshold of -950 HU (RA-950 HU) or the HU corresponding to a threshold at 15% of the lung relative area

145 (Perc15).

Conformance to the requirements of this Profile by all relevant staff and equipment supports the following claims:

Claim 1: Without lung Volume Adjustment (VA), an increase in RA-950* of at least 3.7%, or a decrease in Perc15 of at least 18 HU indicates an increase in the extent of emphysema, with 95% confidence.

Claim 2: With lung VA, a decrease in Perc15 of at least 11 HU, is required for detection of an increase in the extent of emphysema, with 95% confidence.

*Note that in some CT scanners truncation at -1024 HU biases RA-950 values that are near zero; Such cases where low lung density lung values are completely absent, or nearly so, the HU values are not expected to be normally distributed, and the 95% range for the RA-950 measure is useful only as an empirical figure of merit for guiding interpretation of change.

Discussion

The confidence intervals defined for the claims, e.g. -18 and +18 HU in Claim 1 for the Perc15 measure, define boundaries that can be thought of as "error bars" or "noise" around the measurement of lung density. If one measures change within this range, one cannot be certain

that there has really been a change. However, if lung density changes beyond these limits, one can be 95% confident there has been a true change in lung density, and the observed difference is not just measurement variability. Examples of scenarios for measures that are considered within the variability expected, and thus not a true change, vs two measures that exceed the threshold for a true change are illustrated in **Figure 1**.

Note that this does not address the biological significance of the change, just the likelihood that the measured change is real. Once a real change has been identified, the magnitude of the change can be expressed as a 95% CI. For a protocol without VA and given a measured change of Δx HU in Perc15, the true change is expected to lie in the interval [$\Delta x - 18$ HU, $\Delta x + 18$ HU] with 95% confidence; and for a measured change of $\Delta y\%$ in RA-950, the true change is expected to lie in the interval [$\Delta y - 3.7\%$, $\Delta y + 3.7\%$] with 95% confidence. For a protocol with VA, repeatability is improved such that given a measurement of Δx HU in Perc15, the true change is expected to lie in the interval [$\Delta x - 11$ HU, $\Delta x + 11$ HU] with 95% confidence. More detail on how these limits are calculated is described below and provided in Appendix B. However, it bears emphasis that VA should not be thought of as an alternative to breath-hold coaching and control. Consistent breath-hold coaching followed by simple visual inspection (e.g. by watching the subject's chest wall through the scan room window or by camera) to confirm chest inflation is required to meet the claims. VA is a method to further improve repeatability beyond what can be achieved with prospective breath-hold coaching and control.

These claims are based on estimates of the repeatability coefficient (RC) of the RA-950 HU and the Perc15 measured from the histogram of both lungs after segmentation of the thoracic cavity and removal of blood vessels and airways as described in more detail in Section 3.8. The repeatability coefficient (RC) is defined as $1.96 \sqrt{2}$ wSD, where wSD is the within-subject standard deviation. Specific equations used to calculate the expected range for RA-950 HU and Perc15 are defined in Appendix C. The claim assumes that there is negligible proportional bias in the measurements (i.e. bias < 5% of the measurement), and is supported by a meta-analysis (Appendix B) of studies conducted at the same site using the same scan protocol and CT scanner make and model. Further adjustment to remove bias is required when scanning subjects longitudinally on different CT scanner makes and models. Future versions of the Profile that seek to harmonize CT number or HU across different scanner makes and models may address this limitation, but this remains an open issue.

Volume adjustment (VA) refers to techniques to correct for differences in lung inflation volume between time points. The literature has noted that differences in lung inflation volume are present in longitudinal studies and thus repeatability is improved using some type of VA as discussed in more detail below and in Appendix B. There are separate claims for without VA and with VA to reflect the narrower 95% confidence interval with VA. For RA-950, only 2 repeatability studies were available, which was insufficient to support a meta-analysis to inform the impact of VA on the claim for the RA-950 metric. For the studies supporting the stated claims, the method of VA varied. Because more advanced techniques for VA continue to emerge, this document does not intend to suggest any particular model or method for VA. Appendix B provides a description of VA methods that fit the selection criteria. That said,

achieving consistent lung inflation volume through consistent breath-hold coaching and communication as described in Section 3.3.1 is required (and one of the most underappreciated procedures in the workflow) to achieve the claims ^{1,2}. To further guide the various stakeholders interested in quantitative lung density measures using CT, we include protocols in Appendix D that meet or exceed the stated claims.

Clinical interpretation with respect to the magnitude of true change:

Measurement of whole lung parenchymal lung density with CT has been used for several decades in clinical research as a marker of emphysema. For repeat CT examinations that are performed primarily for emphysema quantification, a change in RA-950 HU of $\geq 3.7\%$ of the normalized lung volume, or a change in Perc15 of ≥ 11 HU can be accepted as indicative of a true change in the extent of emphysema, with 95% confidence. Both of these measures reflect specific thresholds of the histogram of lung densities in HU and imply loss of lung tissue based on a combination of comparisons to microscopic histology and associations with known measures of whole lung function. The committee recognizes that these limits of variability are substantially greater than the average change in lung density identified in individual subjects with emphysema. In untreated subjects with alpha-1 anti-trypsin deficiency, the average annual decline in 15th percentile lung density (adjusted for lung volumes) is about 2.2 g/L per year (corresponding to 2.2 HU) ^{3,4}. In cigarette smokers with COPD, the average annual decline in lung density is about 1.1 g/L, or 1.1 HU ⁵. Given this discrepancy, the chief value of measuring change in lung attenuation will be in research cohorts in clinical trials; changes measured in individual subjects are unlikely to exceed the limits of variability. It is hoped that in the future, with greater adherence to this profile, the limits of variability can be narrowed to be more compatible with the changes expected in individual subjects.

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

The performance values in the reported claims reflect the likely impact of variations permitted by this Profile since the meta-analysis was based on studies that incorporated variable methods of CT reconstruction, image analysis, and volume adjustment. The Profile thus allows for the possibility of using variable approaches to attaining lung inflation volume, CT scanner protocol and analysis tools. However, in its current form the Profile does not permit different compliant actors (specifically, acquisition device and image analysis software) be used for both exams of a patient. Again, future versions of the Profile that seek to harmonize CT number or HU across different scanner makes and models could potentially relax this requirement, but this remains an open issue.

3. Profile Activities

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.

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

Table 1: Actors and Required Activities

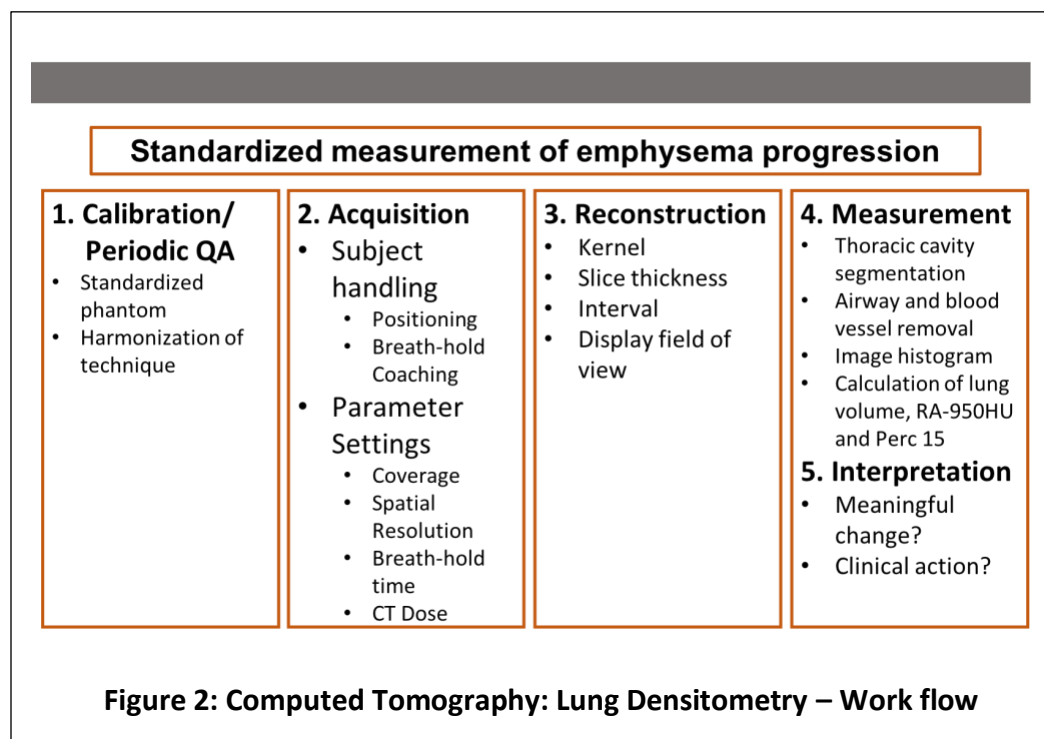
Actor	Activity	Section
Acquisition Device	Scanner Qualification	3.1.
	Periodic QA	3.3.
	Subject Handling	3.5.
	Image Data Acquisition	3.6.
Physicist/Technologist	Scanner Qualification	3.1.
	Staff Qualification	3.2
	Image QA	3.8.
Technologist	Subject Handling	3.5.
	Image QA	3.8.
	Image Data Acquisition	3.6.
	Image Data Reconstruction	3.7.
	Image Distribution	3.9.
Clinician	Subject Handling	3.5.
	Image QA	3.8.
	Image Interpretation	3.11.
Reconstruction Software	Image Data Reconstruction	3.7.
Image Analysis Software	Image Analysis	3.10.
Clinician and Statistician	Image Interpretation	3.11

The requirements in this Profile do not codify a Standard of Care; they only provide guidance

250 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 users of this profile decide to handle deviations for analysis purposes is entirely up to them.

The sequencing of the Activities specified in this Profile are shown in **Figure 2**.

255



3.1. Scanner Qualification

This activity describes performance assessment, calibration or standardization, and validations of equipment that are necessary to reliably meet the Profile Claim.

260 3.1.1 DISCUSSION

These specifications are defined based on groundwork projects from vendor round 1 and 2 studies using the COPDGene Phantom (See “Assessment Procedures”, **Figure 3**) and the QIBA-SRM phantoms⁶.

265 Differences in scanner beam characteristics and calibrations by manufacturer and model are likely sources of systematic variation. At present the profile requires CT baseline and follow-up scans be acquired using the same scanner make and model to meet the claims. A method to empirically correct to a common reference scanner for multi-center studies has been shown to reduce bias and improve precision of quantitative CT measures in phantom testing ⁶, although

this is not strictly necessary to meet the current claims.

270 A model-based correction method based on the composition, and the certified physical
densities of the constituent materials, of the QIBA-SRM phantom⁶ has also shown promise as
an absolute correction method (i.e. standardization). The standardization scheme to be used
across scanner platforms is currently being devised as part of active groundwork and remains
an “open issue” for this version of the Profile. The goal of that work is to eliminate scanner
275 dependent parameters and compare the results of the true material properties such as electron
density across baseline and follow-up scans that can occur potentially with different scanner
makes and models.

Initial qualification of a scanner involves verification that the equipment complies with
specifications described in Table 3.1.2. Subsequent qualification of a scanner for evaluation of
280 longitudinal change in lung density requires calibration of Hounsfield Unit (HU) values to
improve precision and reduce bias. Modern scanners can achieve sub-HU standard deviations
for intra-scanner repeat scans in the lung density region. It is desirable to confine the variations
from different measurement systems to within 1 HU using an adequate test object.

Because of the multitude of software programs used by different commercial, open-source, and
285 academic research labs, an evaluation is warranted to determine the degree of variation, if any,
that different segmentation software applications have on the proposed lung density
measurements used in this profile. Consensus repeatability compiled from commercial and
academic analysis software for RA-950 and Perc15 from a common reference data set (made
available on the QIDW website) are summarized in Section 4.3 and tabulated in more detail in
290 **Appendix E.**

3.1.2 SPECIFICATION

Parameter	Actor	Requirement
Sample Protocol	Acquisition Device	Shall prepare a sample protocol conformant with Section 3.4.2 "Protocol Design Specification"
Noise Performance	Acquisition Device	Shall demonstrate noise bias is $\leq \pm 1$ HU and standard deviation is ≤ 20 HU for lung equivalent foam (approximately -850 HU). See 4.1.2 Assessment Procedure: Voxel Noise and Noise Power Spectrum
In-plane spatial resolution	Acquisition Device	Shall demonstrate a Full-width at half-maximum (FWHM) ≤ 1.0 mm as described in Section 4.1.3.

Through-plane spatial resolution	Acquisition Device	Shall demonstrate a slice sensitivity profile with FWHM ≤ 1.0 mm as described in Section 4.1.4.
Edge Enhancement	Acquisition Device	Shall demonstrate an edge enhancement $\leq 3\%$ for the edge response function as described in Section 4.1.3.
Acquisition speed	Acquisition Device	Shall be capable of whole lung coverage in < 10 s (i.e. a table feed of 3 cm/s or greater).
CT radiation exposure	Acquisition Device	As prescribed in Section 3.4.2.
Measured HU (Bias)	Acquisition Device	Shall demonstrate a mean measured HU of -1000 HU ± 6 HU for inside air (within phantom), and 0 HU ± 6 HU for water (within phantom)*
HU Stability (Repeatability)	Acquisition Device	Shall demonstrate a standard deviation of ≤ 1 HU for inside air (within phantom), lung equivalent foam (within phantom), and water (within phantom) measured across N=5 acquisitions
Lung Density Analysis	Image Analysis Software	Shall be able to calculate and output for the whole lung: <ul style="list-style-type: none"> • RA-950 HU • Perc15 • Lung Density Histogram • Total Lung Volume
Reproducibility of Analysis Software	Image Analysis Software	Shall calculate and output the whole lung the RA-950, Perc15 and Lung Density Histogram: <ul style="list-style-type: none"> • The same analysis software shall be identical for each longitudinal time point measured, and • Above measures shall be deterministic (yield identical results each time the software analysis is applied to the same patient data set) and therefore add no additional variance to the measurement.

		See section 4.2 for more detail on procedures
*This specification is only expected to apply in the COPD Gene, or similarly designed, phantom where uniform low density regions are provided. Both systemic and stochastic noise sources in patient data are expected to introduce bias and reduce precision. This additional variance is accounted for by the repeatability meta-analysis in patient studies used to define the Claims.		

295

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.

300 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.

305 3.2.2 SPECIFICATION

Parameter	Actor	Specification
Qualification	Technologist	Shall be a trained radiation technologist as defined by ARRT.
Qualification	Physicist	Shall be a Qualified Medical Physicist (QMP) as defined by AAPM.

3.3. Periodic QA

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

3.3.1 DISCUSSION

Additional action may be needed if equipment no longer passes performance assessment.

3.3.2 SPECIFICATION

315

Parameter	Actor	Requirement
Monthly QA	Physicist	Shall evaluate the following parameters for each conformant acquisition device at least monthly or after equipment service that may alter its performance.
Re-establishing Standardization	Physicist	Shall, if the acquisition device fails Monthly QA, repeat Product Qualification (See 3.1.2) to re-establish standardization.
Scanner Calibration	Physicist or Technologist	Shall assess the current CT conformance for HU value and standard deviation, encompassing scanning the COPDGene phantom, and analyzing the images to prove conformance on a quarterly basis.
	Acquisition Device	Shall meet the specifications in Table 3.1.2.
HU Stability	Physicist	Shall confirm the longitudinal difference (i.e. drift) between subsequent monthly scans and baseline for air (within phantom) and water (within phantom) falls within ± 4 HU

3.4. Protocol Design

320 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.

325 Generalizable image quality specifications are favored over narrow pre-defined parameter settings to allow flexibility in developing and supporting quantitative density measures within a specified CT dose. Therefore, multiple possible parameter settings are allowed.

330 The assessment procedures for spatial resolution and edge enhancement are performed using an appropriate test object (e.g. ACR or CATphan) to estimate the point or edge response function, and slice sensitivity profile ⁷⁻¹⁰ as described in Section 4.1.3 and 4.1.4. Similar assessment procedures for HU bias and repeatability (Section 4.1.1), and voxel noise (Section 4.1.2) require a COPDGene phantom (See Section 4: “Assessment Procedures”, **Figure 3**) containing foam standards with lung equivalent foam densities. A representative example of a phantom (e.g. COPDGene phantom or equivalent) meeting or exceeding this standard is shown in Figure 3 and described further in Section 4.

335 This approach is intended to enable different vendor architectures and reconstruction algorithms to meet the desired quantitative measurement standards while allowing flexibility to readily adapt protocols as CT systems continue to evolve.

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.

340 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.

345 The purpose of the minimum **scan duration** requirement is to permit acquisition of the lungs in a single breath-hold, thereby preventing respiratory motion artifacts or anatomic gaps between breath-holds.

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

350 **Total Collimation Width** (defined as the total nominal beam width, $N \times T$, for example 64 detectors \times 1.25mm thick) 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.

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

360 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.

365 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, but it is strongly recommended that the CTDIvol be targeted to 3 mGy for an average sized patient (i.e. 75 kg) with the amount of radiation adjusted based on patient size and shape according to manufacturer. CT radiation dose in the chest is expected to vary by approximately $\pm 18\%$ for subject weight between 50-100 kg¹¹, which is acceptable for the longitudinal claim but may be a source of additional variability for comparisons across different scanner makes and models, which is therefore omitted from the current claims of this profile (see "AEC" Open Issue).

370 Image reconstruction is modeled as a separate Activity in the QIBA Profile. Although it is closely

375 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.

380 Many reconstruction parameters can have direct or indirect effects on the lung density
histogram used for computing parenchymal density measures. To reduce this potential source
of variance, all efforts should be made to match acquisition and reconstruction parameters as
with the baseline.

385 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 as described in Section
4.2. The shape of the reconstruction kernel, or modulation transfer function (MTF), alters both
the spatial resolution and noise characteristics of the image ^{10,12}. The reconstruction is a
weighted sample of the structures within the projection. A smoother reconstruction kernel
390 emphasizes larger structures in a projection by increasing their relative weight at the expense
of smaller structures but with the benefit of reducing noise. So there is also an inverse
relationship between spatial resolution and noise that is dependent on choice of reconstruction
kernel, necessitating that the reconstruction kernel be carefully chosen to meet the
specifications in Table 3.1.2 as recapitulated in the context of the human subject protocol in
395 Table 3.4.2. Examples of kernels that would meet the specifications for the major vendors are
described in Section 4.2.

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
400 Noise and Resolution are considered insignificant.

3.4.2 SPECIFICATION

Note that 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
405 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
410 the Physicist actor or equivalent.

Parameter	Actor	Specification	DICOM Tag
Acquisition	Radiologist	Shall prepare a protocol to meet the	

Parameter	Actor	Specification	DICOM Tag
Protocol		specifications in this table. Shall ensure technologists have been trained on the requirements of this profile.	
Total Collimation Width	Technologist	Shall set to Greater than or equal to 16mm.	Total Collimation Width (0018,9307)
Nominal Tomographic Section Thickness	Technologist	Shall set to Less than or equal to 1mm.	Single Collimation Width (0018,9306)
Scan Duration	Technologist	Shall achieve a table speed of at least 4cm per second (in order to cover the full lung within a 10s breath-hold).	Table Speed (0018,9309)
CT Dose	Technologist	≤ 3 mGy CTD _{ivol} for a medium sized (75 kg) subject with the amount of radiation adjusted based on patient size and shape according to manufacturer.	Dose Report
Reconstruction Protocol	Technologist	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	Technologist	Shall set to 1.0 mm or less.	Slice Thickness (0018,0050)
In-plane Resolution	Physicist	Shall validate that the protocol achieves a full width at half maximum (FWHM) of line spread function ≤ 1 mm.	
Through-plane spatial resolution	Physicist	Shall validate that the protocol achieves a slice sensitivity profile with FWHM ≤ 1.0 mm	
Edge Enhancement	Physicist	Shall validate that the protocol achieves a minimum edge enhancement of 3% for the edge response function as described in Section 4.	
Voxel Noise	Physicist	Shall validate that the protocol achieves a standard deviation of voxel noise that is ≤ 20 HU for lung equivalent foam, air and water materials inside a phantom as described in Section 4.	

3.5. Subject Handling

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

3.5.1 DISCUSSION

Subject Positioning

420 Consistent positioning avoids unnecessary variance in attenuation, location of subject within the scan gantry, and changes in anatomical shape due to posture, or body rotation that can affect image quality and consistency of HU value.

Lung Inflation

425 Acquisition parameters have been specified to allow completion of the scan of the whole lung volume in a single breath-hold of less than 10 seconds. Faster scan time can further reduce breath-hold duration and reduce the likelihood of respiratory motion artifacts.

430 Consistency of lung inflation volume is also critical to lung density measures. The specification is to achieve a difference in lung inflation smaller than 10% of baseline lung inflation volume for longitudinal time points with the goal of achieving greater than 90% of predicted TLC at both time points. To achieve consistency of breath-hold it is essential that the technologist perform consistent coaching of the subject before the CT acquisition (so that the subject is prepared for the voice commands while in the scanner).

435 Before the scans are acquired, the coordinator (or trained CT technologist) will review the breathing instructions with the participant and emphasize the importance of following them as closely as possible during the actual imaging of the lungs. In summary, the participant will be instructed to inhale deeply and exhale 3 times and then hold their breath two different ways: with the lungs full of air (TLC scan). The technologist or coordinator should visually confirm that the subject is following the breath-hold coaching as intended (see script below). For example,

440 as individual subjects will vary in their respiratory cycle and compliance with commands, it is important for the technologist or coordinator to give sufficient time for the subject to achieve full inspiration with visual confirmation that this is achieved by watching the subject's chest before CT scanning commences. If additional expiratory scans are performed, please note that the cephalad/caudal coverage of the lungs from apex to base should be adjusted between the

445 TLC and expiratory CT acquisitions to cover the lungs within the limits of the lung apex and base, e.g. no more than 2 cm cephalad to the apical or 5 cm caudal to the basal lung borders.

An additional scout may be acquired between inspiratory and expiratory acquisitions, to prescribe each of the lung volume CT acquisitions separately so as to minimize CT dose to the subject.

450 To extract the desired information from the CT images, it is very important that the breathing

instructions are followed closely. Refer to this publication¹ for further visual and description information on proper breathing instructions.

An example of a breath-hold coaching script is:

- 455 • “Take a deep breath in” (watch the chest to ensure deep breathe is achieved)
- “Let it out” (watch chest to insure exhale is achieved)
- “Take a deep breath in” (watch the chest to insure exhale is achieved and timing of breath cycle for the subject)
- “Let it out”
- 460 • “Now breathe all the way IN, IN, IN...” (watch to confirm timing and inhalation is fully achieved and chest is still)
- “Keep holding your breath – DO NOT BREATHE”
- Visually confirm inspiratory breath-hold by watching subject’s chest and commence CT scan.
- “Breathe and Relax.”

465 3.5.2 SPECIFICATION

Parameter	Actor	Requirement
Subject Positioning	Technologist	Shall place the subject in a supine position, arms positioned comfortably above the head in a head-arm rest with lower legs supported.
Table Height	Technologist	Shall adjust the table height for the mid-axillary plane of the chest to pass through the isocenter.
Subject Alignment	Technologist	Shall position the subject such that the “sagittal laser line” lies along the sternum (e.g. from the suprasternal notch to the xiphoid process).
Scout Scans	Technologist	Shall perform a lateral scout and verify that the mid-axillary plane of the bronchial tree, at the level of the carina, is within 2 cm of iso-center.
		Shall perform an AP (or PA) scout and verify that the subject is correctly centered at horizontal iso-center within 2 cm.
Lung Coverage	Technologist	Shall match the display field of view between time points to insure consistency of spatial resolution. Shall adjust the Cephalad/caudal coverage of the lungs from apex to base between the TLC and expiratory CT acquisitions to cover the lungs within the limits of the lung apex and base, e.g. no more than 2 cm cephalad to the apical or 5 cm caudal to the basal lung borders.

Parameter	Actor	Requirement
Breath-hold Coaching	Technologist	Shall coach the subject on Breath-holding as specified above.
Use of intravenous contrast	Radiologist / Technologist	Intravenous contrast Shall NOT be used.
Use of oral contrast	Radiologist / Technologist	Oral contrast Shall NOT be used.
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.

3.6. Image Data Acquisition

470 This activity describes details of the data acquisition process that are necessary to reliably meet the Profile Claim. It may also include calibrations, performance assessments or validations during acquisition (such as visual confirmation of breath-hold) that are necessary to reliably meet the Profile Claim.

3.6.1 DISCUSSION

475 X-ray CT uses ionizing radiation, and exposure to ionizing radiation increases health risks to the subject. The CT Dose Index Volume (CTDIvol) is used to specify radiation exposure. FDA and international conformance standards require CTDIvol to be available on all CT platforms. The radiation exposure is determined by tube potential, source filtration, tube current-rotation time product, pitch, and total collimation width. The specifications of this profile are designed such that the CTDIvol be less than or equal to 3 mGy for an average-sized subject (75 kg) for each CT scan performed to minimize risk for longitudinal assessment of human subjects.

480

3.6.2 SPECIFICATION

Parameter	Actor	Requirement	DICOM Tag
Acquisition Protocol Selection	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".	
	Technologist	If acquiring a longitudinal time point, shall select	

		a protocol on the same CT scanner with equivalent acquisition and reconstruction parameters to that of the baseline CT scan.	
	Acquisition Device	Qualified Device	Prospective "Quantitative Tag"
Scan Plane	Technologist	Axial / Transverse	Patient Position (0018,5100)
Scout (Topogram, Scanogram)	Technologist	Shall confirm the absence of metal or other artifacts	
Anatomic Coverage	Technologist	Shall ensure the Full Lung, from 2cm above the apex to 5cm below the base, is covered by the scan	
Axial field of view	Technologist	Shall confirm the display field of view is no more than 2 cm outside maximal lung extent. Shall match the display field of view to that of the Baseline scan, if available.	<Confirm DICOM Field> Reconstruction Field of View (0018,9317) Reconstruction Diameter (0018,1100)
CT Dose	Radiologist /Technologist	$\leq 3\text{mGy CTDIvol}$ for a 75kg subject allowing for increased/decreased CT dose adjusted based on patient size and shape according to manufacturer.	CTDIvol (0018,9345) CTDIvol

485

3.7. Image Data Reconstruction

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

3.7.1 DISCUSSION

490 The shape of the reconstruction kernel is a component of the modulation transfer function (MTF) and alters both the spatial resolution and noise characteristics of the image^{10,12}. A smoother reconstruction kernel emphasizes larger structures in a projection by increasing their

495 relative weight at the expense of smaller structures but with the benefit of reducing noise. So there is also an inverse relationship between spatial resolution and noise that is dependent on choice of reconstruction kernel, necessitating that the reconstruction be carefully chosen to meet the specifications in Table 3.1.2 as recapitulated in the context of the human subject protocol in Table 3.4.2. Examples of kernels that would meet the specifications for the major vendors are described in Section 4.2.

500 3.7.2 SPECIFICATION

Parameter	Actor	Requirement	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.	

3.8. Image QA

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

505 3.8.1 DISCUSSION

At the imaging console, subject images will be assessed for:

- Adequate coverage of the thoracic cavity; the lung volume must be fully represented in the field of view.
- Absence of respiratory motion artifact.
- 510 • Appropriate CT dose.
- Appropriate reconstruction algorithm and display field of view as specified.

In conjunction with image analysis, the images will be further reviewed for the above issues, and additionally for the following:

- 515 • Absence of intravenous (IV) contrast.
- (If followup scan) adequacy of lung inflation- no more than 10% difference from baseline scan.
- Specified DICOM fields shall be monitored to confirm adherence to protocol and CT Dose as well as between baseline and followup scans.

520 3.8.2 SPECIFICATION

Parameter	Actor	Requirement
Adequate	Technologist	Shall confirm the lung volume is fully represented in the field of

Parameter	Actor	Requirement
coverage		view
Motion	Technologist	Shall evaluate for respiratory motion (cardiac motion is unavoidable and acceptable) and confirm that the lung parenchyma is sufficiently clear and uncorrupted by motion
Spatial resolution.	Technologist	Shall confirm the image headers (Single Slice Collimation (0018,9306) and Slice Thickness (0018,0050)) indicate the acquired and reconstructed resolutions ≤ 1 mm
CT dose	Technologist	Shall confirm the image header (CTDIvol (0018,9345)) indicates the CTDIvol is "Per protocol."
Conformance to baseline.	Technologist	Shall confirm the Protocol is consistent with the baseline. Shall confirm the FOV (Data Collection Diameter (0018,0090)) are within 10% of baseline except when allowing for specific adjustment to accommodate substantial changes in subject size.
Absence of IV contrast	Image Analysis Software	Shall confirm the absence of IV contrast
Lung Volume	Image Analysis Software	Shall confirm the measured lung volume is within 10% of the baseline.

3.9. Image Distribution

525 Images shall be transmitted in uncompressed DICOM format and according to the anonymization standards approved for the study.

3.10. Image Analysis

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

530 3.10.1 DISCUSSION

535 With the advent of 3D volumetric CT, the field has moved towards full 3D volumetric segmentation of the lung. Given the typically large number of slices in a multi-slice volumetric CT lung scan, (400-600 slices), automated segmentation is a practical necessity. Accurate and reproducible automated segmentation of the lung structures requires combining several segmentation algorithms depending on the level of regional analysis required. For example, reproducible and accurate segmentation (compared to manual analysis) of the right and left lung can be readily performed using optimal thresholding followed by morphological operators and region growing as described by Hu and colleagues¹³. Other methods, such as statistical shape modeling and atlas-based segmentation have been proposed¹⁴.

540

In addition to open source and academic segmentation software, several commercial and prototype commercial packages are becoming available for these analyses (e.g. VIDA, Imbio, Thirona, MeVIS), each with their own proprietary segmentation method. As RA-950 and Perc15 are both straight-forward deterministic computational operations, the primary source of variation in the analysis software lies in the lung volume segmentation mask used to compute the normalizing lung volume, and the removal of the trachea and major airways and vessels and correction of artifacts to a lesser extent.

Due to the multitude of software programs used by different commercial, open-source, and academic research labs, an evaluation is warranted to determine the degree of variation, if any, that different segmentation software applications have on the proposed lung density measurements used in this profile. For example, in longitudinal analysis of CT images that uses different software vendors at different time-points, measurement variability due to the differences in vendor segmentation will be introduced. Consensus reproducibility compiled from various commercial and academic analysis software for RA-950 and Perc15 from a common reference data set (made available on the QIDW website) are summarized in Section 4.2, and tabulated in more detail in Appendix E.

3.10.2 SPECIFICATION

Parameter	Actor	Requirement
Lung Density Analysis	Image Analysis Software	Shall use a consistent lung segmentation procedure including the following steps: <ul style="list-style-type: none"> • Segmentation and removal of central pulmonary blood vessels. • Segmentation and removal of the central airways. • Generation of the image histogram for the remaining lung parenchymal tissues • Shall be identical for each longitudinal time point measured, and • Shall be deterministic (yield identical results each time the software analysis is applied to the same patient data set) and therefore add no additional variance to the measurement. • See section 4.2 for more detail on procedures.
	Image Analysis Software	Shall calculate and output the whole lung RA-950, Perc15 and Lung Density Histogram: <ul style="list-style-type: none"> • The same analysis software shall be identical for each

Parameter	Actor	Requirement
		<p>longitudinal time point measured, and</p> <ul style="list-style-type: none"> • Above measures shall be deterministic (yield identical results each time the software analysis is applied to the same patient data set) and therefore add no additional variance to the measurement. • See section 4.2 for more detail on procedures

560 **3.11. Image Interpretation**

This activity describes criteria and procedures related to clinically interpreting the measurements and images that are necessary to reliably meet the Profile Claim.

3.11.1 DISCUSSION

565 Measured changes in lung attenuation in individuals may be compared with the previously published mean changes; for example the mean change in volume adjusted lung density in untreated subjects with Alpha-1 antitrypsin deficiency is about 2.2 g/L/year³, and the mean change in cigarette smokers with COPD is about 1.1 g/L/year⁵. Important potential biological confounders in measurement of lung attenuation should also be considered. These would include significant changes in inspiratory lung volume, presence of other significant lung diseases on baseline or followup scans (e.g. pneumonia, interstitial lung disease), intervening surgery, and change in smoking status. Importantly, smoking cessation decreases Perc15 lung attenuation by a mean of 4.9 HU, simulating progression of emphysema¹⁵.

570

3.11.2 SPECIFICATION

Parameter	Actor	Requirement
Real Change?	Clinician or Statistician	Determine if disease progressed or is stable.
Magnitude of Change	Clinician or Statistician	Determine if magnitude of the change <i>is</i> or <i>is not</i> significant with respect to the hypothesis of the proposed study.

575

4. Assessment Procedures

Most of the requirements described in Section 3 can be assessed for conformance by direct observation, however some of the performance-oriented requirements are assessed using a procedure. Specific assessment procedures that are required, or need further clarification, are defined in subsections below and the subsection is referenced not in numerical order, but from the corresponding requirement in Section 3.

Below are specific examples of assessment procedures that meet the standards of this profile. In each case a specific measurement procedure is outlined using the COPDGene phantom (**Figure 3**) that is sufficient to meet the minimum standards for scanner qualification (Section 3.1) and periodic quality assurance (Section 3.2) and is published and commercially available (the COPDGene phantom, The Phantom Laboratory - <https://www.phantomlab.com/catphans-copd>)^{16,17}. For more advanced studies in which standardization is required across a network of sites, an ideal reference object would include a series of at least five foam standards whose density is in the range of lung parenchyma (64-321 kg/m³) that have been calibrated to their true densities. Such a phantom is introduced and described in⁶ and has been shown to establish a HU-electron density relationship for a given scanner and protocol. Related methods¹⁸ may also be considered for harmonization, although this is not strictly necessary to meet the longitudinal claims.

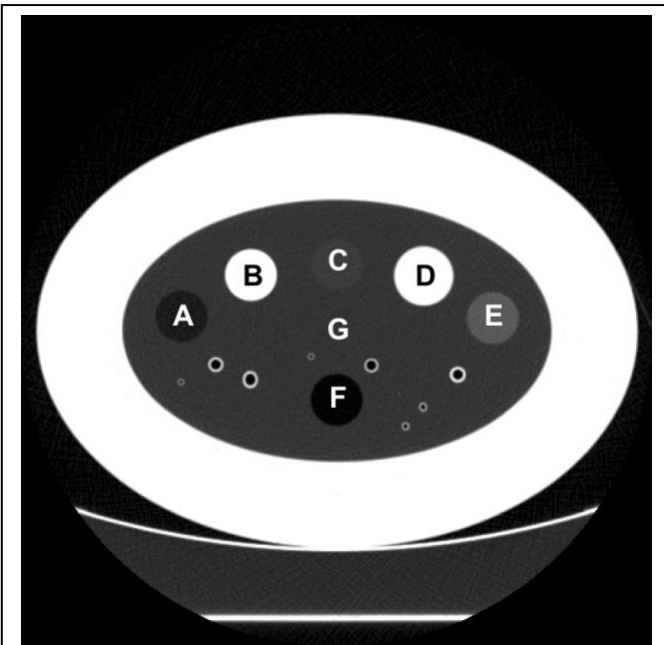


Figure 3. CT slice image of the “COPD Gene” lung phantom used in this study. The phantom consists of (A) 4 lb/ft³, (B) acrylic, (C) 12 lb/ft³, (D) water, (E) 20 lb/ft³, (F) air, and (G) a larger oval shaped lung density equivalent foam surrounded by an outer chest wall equivalent uniformity ring.

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

The test procedure described here is based on the use of conventional filtered backprojection reconstruction methods; extreme care must be taken when iterative reconstruction methods are used as their use may invalidate some of the assumptions inherent in this method and are considered open issues for the present status of this Profile.

Several CT protocols are also given in a link within Appendix D that are derived from the proposed procedures and would meet the claims of this profile. Acquisition parameters have been specified to allow

completion of the scan of the whole lung volume in a single breath-hold of less than 10 seconds. Faster scan time can further reduce breath-hold duration and reduce the likelihood of respiratory motion artifacts. This procedure is recommended to be used by a physicist at the vendor origin or at an imaging site, to insure product qualification for a CT scanner in terms of linearity and repeatability of the measurement of CT number for air, lung equivalent foam, and water standards.

For the minimum standard, the following procedure can be followed for image acquisition:

1. 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.
2. It is expected that scout (topogram, scanigram, etc.) images will be initially obtained to optimize positioning and coverage of an imaging phantom.
3. The assessor shall then scan a phantom containing regions of uniform low density such as the COPD Gene phantom^{16,17}. The phantom shall be placed at the isocenter of the scanner. The acquisition protocol and reconstruction parameters shall conform to this Profile (See Section 3.6.2 and 3.7.2). The same protocol and parameters shall be used when performing the assessments in 4.1.1, 4.1.2 and 4.1.3.

4.1.1 Assessment Procedure: HU Bias and Repeatability

The procedure for measuring water and air density in the COPD Gene phantom is performed as follows:

1. As for any phantom calibration in CT, it is critical to position the phantom at the center of gantry rotation and in alignment with the axial scan plane such that the standards are minimally affected by differences in magnification and resolution. This can be achieved by using fiducial markers on the phantom and the alignment system of the scanner itself. Scanning should be performed for N = 5 repeated realizations (i.e. repeated acquisitions of the phantom) in order to measure the noise and standard deviation of mean HU values in the procedures below and according to the specifications in Table 3.1.2.
2. After the scan is performed, the assessor shall place an approximately cylindrical region of interest (ROI) within a central axial slice of the phantom to include the uniform volume of the material to be measured (e.g. inside water and air, Figure 3D and F, respectively) positioned such that partial volume effects will NOT impact the measurement. An ROI shall be drawn to avoid the outer 2 mm of the cylindrical circumference but containing at least 3,000 voxels for these measurements. Standard commercial and opens source software packages that are sufficient for ROI placement and measurement are published¹⁹, [e.g. <http://airwayinspector.acil-bwh.org/>].
3. The assessor shall record the values reported for the ROI mean and standard deviation

of the HU and insure they meet the specifications in Table 3.1.2. Specifications are restated here for convenience and include:

655

- a. 1000 HU \pm 6 HU for inside air (within phantom), and 0 HU \pm 6 HU for water (within phantom), and
- b. for repeated scans across N=5 acquisitions, a standard deviation of \leq 1 HU for inside air (within phantom), lung equivalent foam (within phantom), and water (within phantom).

660

4.1.2 Assessment Procedure: Voxel Noise and Noise Power Spectrum

This procedure can be used by a vendor or an imaging site to assess voxel noise. For voxel noise assessment, use the same scanning procedure and images acquired as in 4.1.1 above.

The procedure for measuring voxel noise is performed as follows:

665

1. Subtract two unique realizations of the phantom and place a cylindrical ROI in the center of the phantom in a uniform regions corresponding to the lung density equivalent foam insert regions (Figure 3G). ROI placement and size should meet the minimum standards in 4.1.1 above.
2. The assessor shall record the values reported for the ROI mean and standard deviation and insure they meet the specifications for noise bias and standard deviation specified in Table 3.1.2. Specifications are restated here for convenience and include:
 - a. noise bias in the subtracted image is $\leq \pm$ 1 HU, and
 - b. the standard deviation of the noise is \leq 20 HU.

670

675

Given the proprietary nature of CT reconstruction kernels, some care in comparing reconstruction kernels, or “algorithms”, used for reconstruction across vendors is warranted. The vendor shall specify that the reconstruction kernel for the proposed protocol design (Section 3.4 and Table 3.4.2) is matched as best as possible to the spatial frequency dependence of the noise behavior, i.e. the noise power spectra (NPS), of typical smooth kernels such as the GE Standard, Siemens B31f or B34f, Philips B, and Cannon FC17 kernels. Methods for calculating the NPS are published^{10,12}. For multi-center protocols that use different CT vendors in particular, calculation of the NPS and certification by the vendor that all reconstruction kernels are well matched with respect to their NPS is critical. Other pairings and novel reconstruction kernels or acceptable assuming they meet the specifications outlined in Section 3.1 and 3.4, are sufficiently smooth, and are matched throughout the network.

680

685

4.1.3 Assessment Procedure: In-Plane, Through-Plane (Z-axis) Spatial Resolution and Edge Enhancement

The vendor shall confirm that the slice profile, point spread and edge response functions, PSF and ERF respectively, meet the standards specified for scanner qualification (Section 3.1 and Table 3.1.2) and are within adequate range for protocol design (Section 3.4 and Table 3.4.2).

690 The derived point response function shall meet the specifications outlined for spatial resolution
in Table 3.1.2. Specifications are restated here for convenience and include:
c. The point response function shall demonstrate a Full-width at half-maximum
(FWHM) ≤ 1.0 mm.

695 Procedures for assessing PSF and ERF has been published using the derivative of the edge
response function (ERF) measured from the oversampled outside edge of, for example, the
phantom as in the reference by Judy ⁷, or using standard CT performance phantoms such as the
ACR phantom or CATphan.

700 Using either phantom, the procedure for measuring Edge Enhancement (EE) shall be performed
as follows:

1. EE shall be derived from the edge response function according to the equation:

$$EE (\%) = 100 \left(\frac{EE_m}{EE_r} - 1 \right),$$

705 Where EE_m is the maximum observed contrast along the ERF and EE_r is the reference
value calculated as the mean HU value within a uniform region of the edge material (e.g.
the outer chest wall equivalent uniformity ring). The derived response functions shall
meet the specifications outlined for edge response in Table 3.1.2. Specifications are
restated here for convenience and include:

- a. Maximum edge enhancement $\leq 3\%$.

Using the CATphan or similar, the procedure for measuring through-plane resolution (slice
profile) shall be performed as follows:

- 715 1. The through-plane resolution shall be assessed using the slice sensitivity profile using
methods described in Fuchs et al. ⁸ and with standard procedures with CT performance
phantoms such as the ACR or CATphan phantoms. The derived slice sensitivity profile
shall meet the specifications outlined for through-plane resolution in Table 3.1.2.
Specifications are restated here for convenience and include:

- a. Shall demonstrate a Full-width at half-maximum (FWHM) ≤ 1.0 mm.

4.2. Assessment Procedure: Reproducibility of Image Analysis Software across Various Vendors

The analysis software:

- 725 1. Shall be identical for each longitudinal time point measured, and
2. Shall be deterministic (yield identical results each time the software analysis is
applied to the same patient data set) and therefore add no additional variance to
the measurement.

730 As part of the profile development, the lung density biomarker committee collected and tested
analysis software performance on a reference data set of COPD patient data with a range of
severities. The reference data set is made available for download upon request at the
quantitative imaging data warehouse ([QIDW](#)). The reference data set consists of 50 cases with
735 varying degrees of airflow limitation, ranging from never-smokers with normal pulmonary
function (n=10) to GOLD IV COPD (n=10 cases in each GOLD group), and including both
conventional (~6 mGy average CTDIvol) and reduced dose (~3 mGy average CTDIvol) CT data
sets from the same subjects. To aid developers, the performances of lung segmentation, RA-
950 and Perc15 for the academic and commercial software vendors using the RDC were
740 compared in Appendix E for repeated measurements of the same subjects from the reference
data set. With this data it is possible to compare performance of the RDC for a new software
analysis tool. One standard for reproducibility performance of a new software analysis tool
compared to existing analysis tools would be that conformal measurement of the RDC for TLC,
RA-950 and Perc15 should be less than 0.31L, 1.2% and 1.7HU, respectively (Appendix E:
Median values from Table 2). One process for calculating RDC to compare to the performance
of other software analysis tools is described in Appendix F.

745 **5. Conformance**

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

750 To support an activity, the actor shall conform to the requirements (indicated by “shall
language”) listed in the Specifications table of the activity. Each activity has a dedicated
subsection in Section 3. For convenience, the Specification table requirements have been
duplicated and regrouped by actor in the form of a checklist in Appendix E.

Some requirements reference a specific assessment procedure in section 4 that shall be used to
assess conformance to that requirement.

755 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.

Formal claims of conformance by the organization responsible for an Actor shall be in the form
of a published QIBA Conformance Statement.

760 Vendors publishing a QIBA Conformance Statement shall provide a set of “Model-specific
Parameters” (as shown in Appendix D) describing how their product was configured to achieve
conformance. Vendors shall also provide access or describe the characteristics of the test set
used for conformance testing.

References

765 1. Newell JD, Jr., Sieren J, Hoffman EA. Development of quantitative computed
tomography lung protocols. *J Thorac Imaging*. 2013;28(5):266-271.

2. Sieren JP, Newell JD, Jr., Barr RG, et al. SPIROMICS Protocol for Multicenter Quantitative Computed Tomography to Phenotype the Lungs. *Am J Respir Crit Care Med*. 2016;194(7):794-806.
3. Dirksen A, Piitulainen E, Parr DG, et al. Exploring the role of CT densitometry: a randomised study of augmentation therapy in alpha1-antitrypsin deficiency. *Eur Respir J*. 2009;33(6):1345-1353.
- 770 4. Chapman KR, Burdon JG, Piitulainen E, et al. Intravenous augmentation treatment and lung density in severe alpha1 antitrypsin deficiency (RAPID): a randomised, double-blind, placebo-controlled trial. *Lancet*. 2015;386(9991):360-368.
- 775 5. Coxson HO, Dirksen A, Edwards LD, et al. The presence and progression of emphysema in COPD as determined by CT scanning and biomarker expression: a prospective analysis from the ECLIPSE study. *Lancet Respir Med*. 2013;1(2):129-136.
6. Chen-Mayer HH, Fuld MK, Hoppel B, et al. Standardizing CT lung density measure across scanner manufacturers. *Med Phys*. 2017;44(3):974-985.
- 780 7. Judy PF. The line spread function and modulation transfer function of a computed tomographic scanner. *Med Phys*. 1976;3(4):233-236.
8. Fuchs T, Krause J, Schaller S, Flohr T, Kalender WA. Spiral interpolation algorithms for multislice spiral CT--part II: measurement and evaluation of slice sensitivity profiles and noise at a clinical multislice system. *IEEE Trans Med Imaging*. 2000;19(9):835-847.
- 785 9. Zhang J, Bruesewitz MR, Bartholmai BJ, McCollough CH. Selection of appropriate computed tomographic image reconstruction algorithms for a quantitative multicenter trial of diffuse lung disease. *J Comput Assist Tomogr*. 2008;32(2):233-237.
10. Friedman SN, Fung GS, Siewerdsen JH, Tsui BM. A simple approach to measure computed tomography (CT) modulation transfer function (MTF) and noise-power spectrum (NPS) using the American College of Radiology (ACR) accreditation phantom. *Med Phys*. 2013;40(5):051907.
- 790 11. Huda W, Sterzik A, Tipnis S, Schoepf UJ. Organ doses to adult patients for chest CT. *Med Phys*. 2010;37(2):842-847.
12. Solomon J, Samei E. Quantum noise properties of CT images with anatomical textured backgrounds across reconstruction algorithms: FBP and SAFIRE. *Med Phys*. 2014;41(9):091908.
- 795 13. Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans Med Imaging*. 2001;20(6):490-498.
14. Sluimer I, Schilham A, Prokop M, van Ginneken B. Computer analysis of computed tomography scans of the lung: a survey. *IEEE Trans Med Imaging*. 2006;25(4):385-405.
- 800 15. Shaker SB, Stavngaard T, Laursen LC, Stoel BC, Dirksen A. Rapid fall in lung density following smoking cessation in COPD. *COPD*. 2011;8(1):2-7.
16. Sieren JP, Newell JD, Judy PF, et al. Reference standard and statistical model for intersite and temporal comparisons of CT attenuation in a multicenter quantitative lung study. *Med Phys*. 2012;39(9):5757-5767.
- 805 17. Guo J, Wang C, Chan KS, et al. A controlled statistical study to assess measurement variability as a function of test object position and configuration for automated surveillance in a multicenter longitudinal COPD study (SPIROMICS). *Med Phys*. 2016;43(5):2598.

- 810 18. Martinez LC, Calzado A, Rodriguez C, Gilarranz R, Manzanas MJ. A parametrization of the CT number of a substance and its use for stoichiometric calibration. *Phys Med.* 2012;28(1):33-42.
19. J. Guo, M. K. Fuld, S. K. Alford, J. M. Reinhardt, Hoffman. EA. Pulmonary Analysis Software Suite 9.0: Integrating quantitative measures of function with structural analyses. Paper presented at: First International Workshop on Pulmonary Image Analysis2008; New York.
- 815

Appendices

820 **Appendix A: Acknowledgements and Attributions**

This Profile is a joint effort of the QIBA Lung Density Biomarker Committee* under the supervision of Phil Judy, Heather Chen-Mayer, Sean B. Fain, David Lynch, and Matthew K. Fuld. Discussions and feedback from Nancy Obuchowski greatly improved the statistical methods used in the meta-analysis. Discussions and editorial comments from Greg Kinney, and Ella
825 Kazerooni improved the presentation to better support the profile.

*Committee members: Sean B. Fain, David Lynch, Matthew Fuld, Philip Judy, Heather Chen-Mayer, Stephen Humphries, Bernice Hoppel, Charles Hatt, Miranda Kirby, Jared Sieren, and Amin Motahari.

830 **Appendix B: Background Information**

1. Introduction

This Appendix reports details on how the specific claims were derived at the time of conclusion of literature reviews (8/4/2014) on the use of computed tomography (CT) measures of lung parenchymal density as a method for estimating severity and progression of emphysema in the
835 lungs. Only whole lung measurements are considered. Regional and lobar measures of emphysema are increasingly being investigated and reported in the literature¹⁻⁶. However, the number of studies using regional measures is currently insufficient to assess emphysema severity, likewise their bias and repeatability for studies of emphysema progression.

Measurement of whole lung parenchymal lung density with CT has been used for several
840 decades as a clinical research marker of emphysema^{3,7-10}, but have not been widely adopted in clinical practice. With the advent of reduced dose lung cancer screening CT examinations, now recommended annually by the U.S. Preventive Services Task Force for 55-80 year olds with a 30 or more pack-year history of smoking, the value of serial emphysema CT measurements based on lung density as important clinical data becomes immediately more relevant¹¹. This will
845 become substantially more important if an effective treatment for emphysema becomes available. If repeat CT examinations were to be performed primarily for emphysema

quantification, reduced dose techniques are important given the life expectancy of patients this could be applied to and the number of examinations they may undergo. Multiple but related measures of parenchymal density have been applied in clinical research, most prominently the relative area (RA), or low attenuation area (LAA), below specific thresholds of the histogram of lung densities in Hounsfield units (HU). Typically thresholds from -970 through -910 HU are used, at inspiratory lung volume coached to total lung capacity (TLC). The most common thresholds used are RA-950 HU, and RA -910 HU. The RAs are expressed in fractions or percentages. A second related class of measures inverts the relative area of lung below a threshold by specifying a single HU value below which a fixed relative lung area falls. For example, common measures of this class include the HU threshold below which 1% (Perc 1) or 15% (Perc15) of the lung area falls, respectively.

Unfortunately only a limited number of studies have compared density measurement in the lungs to micro and macroscopic measures of histology derived from similar regions of diseased lung. Measures of tissue histology in the lungs are the gold standard for measuring severity and progression of emphysema but are necessarily limited themselves. Typically, such studies in human subjects or patients depend on obtaining resected tissue after lung surgery^{7-9,12}. Consequently, only limited agreement exists as to which of the related CT measures of lung parenchymal density is superior for detecting the presence and severity of emphysema. The more common measures have some empirical consensus based on a combination of comparisons to microscopic histology and associations with known measures of whole lung function. This consensus supports RA-950 HU^{8,9,12} and Perc 1¹² as the measures best correlated to microscopic histology, and Perc15⁴ as the measure that has undergone the greatest degree of empirical validation and shown to be highly correlated to lung function^{13,14}. Therefore, RA-950 HU and Perc15, being the most studied and best validated measures in clinical research studies, are hereby recognized as the reference standards and are used to determine the claims for bias and repeatability of lung parenchyma measures in this document.

Some general limitations of both RA-950 HU and Perc15 should be recognized. RA-950 is especially sensitive to noise, which varies with choice of image reconstruction kernel and mA

used for image acquisition. High frequency reconstruction kernels (so-called “hard” algorithms) result in higher absolute RA-950 irrespective of disease severity ¹⁵⁻¹⁷. The Perc15 measure is less sensitive to image noise but is still affected ⁴. Both measures are sensitive to the state of lung inflation or deflation. There is strong underestimation by RA-950 and overestimation by Perc15 (the two measures move in opposite directions) if lung inflation is less than 90% of TLC ¹⁸. This has necessarily focused significant research effort on lung volume adjustment (VA) methods ¹⁹. Lung VA appears to be justified for longitudinal studies where reduction in the limits of agreement in Bland-Altman analyses is substantial, e.g. on the order of 40% (Table 3), after correction using either statistical regression methods ²⁰, or the constant lung mass assumption referred to as the “sponge” model ^{21,22}. It is important to remember that progressive emphysema results in increased lung volumes, and there is some concern that correction for lung volumes may therefore reduce the apparent increase in emphysema. However, in longitudinal studies, measures of emphysema corrected for lung volume have more consistently demonstrated progression than uncorrected values ^{23,24 25}. In Tables 2 and 3 we tabulate the bias and limits of agreement for a subset of longitudinal studies, both before and after performing VA, reported in the above selected studies using linear regression models with fixed and/or random effects. In particular, the sponge model represents the generalized approach to VA applicable to longitudinal studies, i.e. when correction of follow-up lung volume is made relative to a baseline scan. Expected improvements in bias and precision after lung VA are further discussed in Section 3.

2. Study Inclusion and Exclusion Criteria

The studies included for estimating the bias and precision of lung parenchymal density with CT had to meet the following inclusion criteria:

- 1) Publication year of the study after 2005 because CT scanner architectures and stability changed radically when these methods migrated from axial to multi-slice helical platforms and protocols.
- 2) A minimum of 16 slice detectors were used for CT acquisition to limit maximum breath-hold time to ~10 s with 3D whole lung coverage.
- 3) The same or similar CT platform was used for repeated scans.

- 905 4) The Study methods provided sufficient details regarding CT reconstruction and acquisition parameters to verify consistency.
- 5) Sufficient data for analysis was included to conduct Bland-Altman analysis to calculate bias and limits of agreement for one or both of RA-950 and Perc15.
- 6) Subjects were scanned twice or more with less than or equal to a 4 month interval
910 between CT scans with the intent of eliminating the influence of possible disease progression on the bias and precision estimates ²⁶.

Studies were explicitly excluded if:

- 1) Repeatability data were not included.
- 2) Parenchymal density was not measured with either RA-950 or Perc15.
- 915 3) The time interval between repeated CT scans exceeded 4 months or other inclusion criteria listed above were not met.

It is inevitable that even for the included studies variability remained, which was mainly in these four areas: CT scanner platform used, Number of subjects, severity of disease, and time interval between CT scans. The studies included are summarized and referenced in Table 1. The
920 density metrics reported in this document are from studies on subjects assumed to be free of disease progression during the short time intervals between baseline and repeat scans. The primary sources of the within-subject variations in the apparent density of the lung parenchyma are attributed primarily to inspiration levels and scanner calibration/measurement error. This approach constrains sources of variation not due to the measurement method. Moreover, at
925 the present time there are few studies that address repeatability with whole lung volume coverage in a reasonable breath-hold (<10 s) even over longer time intervals. There are two major reasons repeatability studies can be excluded: due to scanner architecture ^{21,27-30}, and due to excessive time between scans or inconsistency of methodology ^{14,22,31-33}. One exception to this is the Park et al study ²²that was included as a reference example in an “asymptomatic
930 population” (GOLD stage 0) with a longer time interval between CT scans (8 months), but the subjects were deemed to have “no perceptual changes of disease progression”.

3. Effect of VA on Bias and Precision Claims for RA-950 and Perc15

It is generally recognized that VA is useful for improving the precision of repeat CT measurements of lung density ²⁹. The cross sectional (inter-scanner) variation issue can be addressed by phantom studies ³⁴. In the current document we focus on longitudinal clinical studies from the same site using the same scan protocol, examining the sample mean bias and limits of agreement in the absence of disease progression after VA.

Generally the underlying physiological model-- the sponge model in which the lung mass is assumed to be conserved -- affords the simplicity of an inverse proportionality between lung density and volume. However, this model is not strictly followed in clinical data or even in phantom studies ³⁵. The more common approach is a statistical model that assumes a linear combination of effects that contribute to the density variation in repeat scans. The simplest is the linear fixed effect model in which the change in density metrics (dependent variable) is paired with the change in volume (independent variable), and the linear regression analysis returns a slope and intercept which are then used to correct the density metrics in the repeat scan, such that it may now be compared to the baseline scan free of inspiration related variation. The more advanced model adds random effects to address the individual level of variation, recognizing that the fixed effect model only takes into account the variation at the cohort level. Reference ²³ has a detailed comparison of different models in the study of clinical data in the context of disease progression. For the studies compiled in this document, the method of VA varies, and is reported without any assessment of statistical or clinical merits. Because more advanced techniques for VA continue to emerge ³⁶, this document does not intend to suggest any particular model or method, but rather provides an analysis of the sample studies available that fit the selection criteria. The results of the VA from the included studies are summarized in Table 3. The data, bias and limits of agreement (LOA) before and after VA are plotted in Figure 1.

4. Mean Repeatability Coefficient Obtained by Meta-Analysis

An initial assessment ³⁷ was performed on a subset of data listed in Table 3 with a fixed effect linear model to test whether VA had an effect on the bias and LOAs of the change in Perc15 and RA-950. For Perc15, four studies were included, and the mean bias of -0.168 before VA and 0.090 after VA are statistically insignificant ($p=0.46$). Therefore, it was concluded that VA has

negligible effect on the bias. Due to the lack of homogeneity in LOA exhibited by this set of studies, a more rigorous analysis was performed following the QIBA metrology working group guidelines³⁸. QIBA statistical methods guidelines³⁹ and terminology guidelines⁴⁰ recommend use of the repeatability coefficient (RC) instead of the LOA for the technical performance assessment of repeatability of a quantitative imaging biomarker (QIB). Therefore, the RC, which is a function of the within-subject variance, was first obtained based on the reported bias and LOA values for each study, and the meta-analysis was conducted using the random effects model³⁸. The results of these analyses are summarized in Table 4 and in the Forest plots in Figure 2. The concept of Smallest Real Difference (SRD) is defined by the RC following the conventions of the published literature for test-retest assessments⁴¹⁻⁴³. More details on the RC and the meta-analysis are in the Appendices of this document. The claim language in Section 5 is separated into without and with VA to reflect a narrower 95% confidence limit as a result of VA. For RA-950, only Ref.¹⁵ fits the inclusion criteria, and despite making an exception to also allow the Park et al study²², only 2 studies are insufficient to support a meta-analysis to inform the claim for the RA-950 metric. The committee will monitor the literature and make a more definitive claim when using the RA-950 metric in combination with VA when more data becomes available.

5. Claims:

- For longitudinal studies with reduced-dose CT when monitoring patients who have smoking-induced emphysema:
- i. Without lung VA, an increase in RA-950 of at least 3.7%, or a decrease in Perc15 of at least 18 HU, is required for detection of an increase in the extent of emphysema, with 95% confidence.
 - ii. Without VA, for a measured change of Δx HU in Perc15, one can expect the true change to lie in the interval $[\Delta x - 18 \text{ HU}, \Delta x + 18 \text{ HU}]$ with 95% confidence; for a measured change of $\Delta y\%$ in RA-950, one can expect the true change to lie in the interval $[\Delta y - 3.7\%, \Delta y + 3.7\%]$ with 95% confidence.
 - iii. With lung VA, a decrease in Perc15 of at least 11 HU, is required for detection of an increase in the extent of emphysema, with 95% confidence.*

iv. With VA, from a measured change of Δx HU in Perc15, one can expect the true change to lie in the interval $[\Delta x - 11 \text{ HU}, \Delta x + 11 \text{ HU}]$ with 95% confidence.

* No claim is made for the RA-950 measure with volume correction due to the lack of sufficient data at the time of this report.

Table 1. Summary of parameters for the selected studies used to determine the claims.

Author	Year	Number of Subjects	Disease Severity	Time Interval (mean or median)	Scanner Manufacturer and Model	Density Measures	Reconstruction Kernel	Number of Centers & Vendor Platforms	Tube Current-Time (mAs)	Meets Study Inclusion Criteria? (Reason)
Park ²²	2012	52	GOLD 0	8 mo	Siemens Sensation 16	RA-950, Perc 1, Perc15	B30f	1, 1	40	N (Time Interval) [#]
Chong ¹⁵	2012	44*	Mild COPD (>GOLD 0)	7 days	GE, Siemens, Toshiba	RA-950, Perc15	Bone, B45f, FC51 (high freq.) ^b	3, 3	80-150	Y
Keller ²⁰	2011	105 ⁺	Mild assumed (cancer cohort)	< 4 mo, 78±27 days	GE Lightspeed Ultra (16 detectors)	RA910, Perc15	Bone	1, 1	40	Y
Hochhegger ¹⁷	2011	50* ⁺	No clinical emphysema or lung cancer; >20 pack-year smokers	< 3 mo, 78 days	Phillips Brilliance 64	RA-950	D (Smooth)	1, 1	200	Y
Diciotte ⁴⁴	2011	99* ^o	NCE or lung cancer >20 pack years	3 mo	Siemens Sensation 16	RA-950, Perc15	“Sharp”; 3X3 Gaussian [‡]	1, 1	30	Y [‡]
Gietema ²⁶	2007	157	“Heavy” Smokers	3 mo	Phillips MX800IDT or Brilliance 16p	RA-950	B (moderately soft)	1, 2	30 **	Y**

*Subset of a multi-center, multi-platform study for which two baseline scans were obtained.

⁺Subset of subjects with < 4 months between scans

*⁺Subset of 475 subjects screened for cancer

*^oMILD trial only (16 slice)

** 140 kVp (vs. 120 kVp) for > 80kg weight

[‡]Smoothing operation in image domain

^b High frequency instead of smooth kernel used for reconstruction.

[#]Note that Park et al is retained in subsequent analysis in spite of the longer scan intervals, considering the study subjects were an “asymptomatic population” (GOLD stage 0) and that the subjects had no perceptual changes of disease progression.

Table 2. Summary of repeated measures, bias, and LOA for RA-950 and Perc15 for baseline and repeat scans from the selected studies without volume adjustment.

Author	Year	Lung Volume (ml)		RA-950 (%)				Perc15 (HU)			
		Mean (SD) Time 1	Mean (SD) Time 2	Mean (SD) Time 1	Mean (SD) Time 2	Bias (SD)	Limits of Agreement	Mean (SD) Time 1	Mean (SD) Time 2	Bias (SD)	Limits of Agreement
Park	2012	4784 (908)	4806 (978)	8.7 (5.3)	9.2 (5.1)	0.5 (3.5)	-6.4, 7.4	-933.4 (13.8)	-934.5 (13.7)	-1.1 (10.9)	-22.1, 20.7
Chong	2012	5770 (1540)	5724 (NR)	16.0 (11.7)	NR	0.01 (1.17)	-2.33, 2.35	-946.3 (28.3)	NR (NR)	0.52 (5.29)	-10.07, 11.11
Keller	2011	NR	NR	-	-	-	-	NR	NR	-0.39 (10.1)	-20.13, 19.35
Hochhegger	2011	2578 (584)	2518 (591)	0.53 (0.77)	0.71 (1.19)	0.39 (0.88)	-1.35, 2.15	-	-	-	-
Diciotte	2011	6290 (1220)	6270(1180)	6.2 (6.7)	6.5 (6.8)	0.3 (3.4) [#]	--6.8 , 6.6	-925.2 (22.3)	-925.3 (21.8)	-0.1 (11.5)	-20.1, 25.0
Gietema	2007	6935 (1267)	6945 (1322)	0.17 (NR)	0.08 (NR)	-0.09 (1.19)	-1.3, 1.1	-	-	-	-

NR – “not reported.”

For ‘-’ the measure was not an end-point for the published study.

[#] Diciotte did not report the SD for the RA-950 bias or LOAs; values here are estimated based on the pooled SD of the mean from time 1 and 2, assuming a correlation coefficient being comparable to the one given for the lung volume between time 1 and 2.

Table 3. Comparison of bias and LOA for RA-950 and Perc15 for baseline and repeat scans from the selected studies with VA

Author	Method	RA-950 HU (%)						Perc15 (HU)					
		Without VA		With VA				Without VA		With VA			
		Bias (SD)	LOA	Bias (SD)	LOA	Δ LOA	% Δ	Bias (SD)	LOA	Bias (SD)	LOA	Δ LOA	% Δ
Park	Linear Regression	0.5 (3.5)	-6.4, 7.4	1.1 (1.9)	-3.6, 3,7	-5.2	47	-1.1 (10.9)	-22.1, 20.7	0.3 (6.6)	-12.6, 13.2	-8.5	40
Chong	Linear Regression	0.01 (1.17)	-2.33, 2.35	0.22 (0.83)	-1.43, 1.87	-0.69	29	0.52 (5.29)	-10.7, 11.11	-0.42 (1.82)	-4.05, 3.21	-7.3	67
Keller ¹	Univariate Linear Regression	0.22 (5.19)	-9.95, 10.38	-0.27 (3.25)	-6.63, 6.10	-3.8	37	-0.39 (10.07)	-20.13, 19.35	0.48 (6.86)	-12.97, 13.94	-6.3	32
	Multi-variate Linear Regression			-0.13 (2.99)	-5.98, 5.73	-4.3	42			0.41 (5.67)	-10.7, 11.52	-8.64	44
Hochegger	No VA	-	-	-	-	-	-	-	-	-	-	-	-
Diciotte ²	Linear Regression (log-transformed)	-	-	-	-	-	-	-0.1 (11.5)	-20.1, 25.0	0.0 (8.2)	-14.7, 17.3	-6.51	29
Gietema	No VA	-	-	-	-	-	-	-	-	-	-	-	-

For '-' the measure was not an end-point for the published study.

Shaded cells are not included due to missing data or if for RA-910 HU (as in Keller et al).

¹Keller et al shows data from RA -910 HU (omitted from Table 2) and Perc15 (as reported in Table 2).

²linear regression of the change in log-transformed Perc15 vs change in log-transformed TLV.

Table 4. Summary of the results of meta-analyses for the repeatability coefficient (RC).

Summary table for meta-analysis of RC		#studies	Summary Estimate of RC (Weighted Mean)	SE	95% CI (2-sided)	Lower bound	Upper bound
Perc15 (HU)	without VA	4	18.4	3.0	5.9	12.6	24.3
	with VA	4	11.4	3.5	6.8	4.6	18.3
RA-950 (%)	without VA	5	3.7	0.7	1.4	2.3	5.1

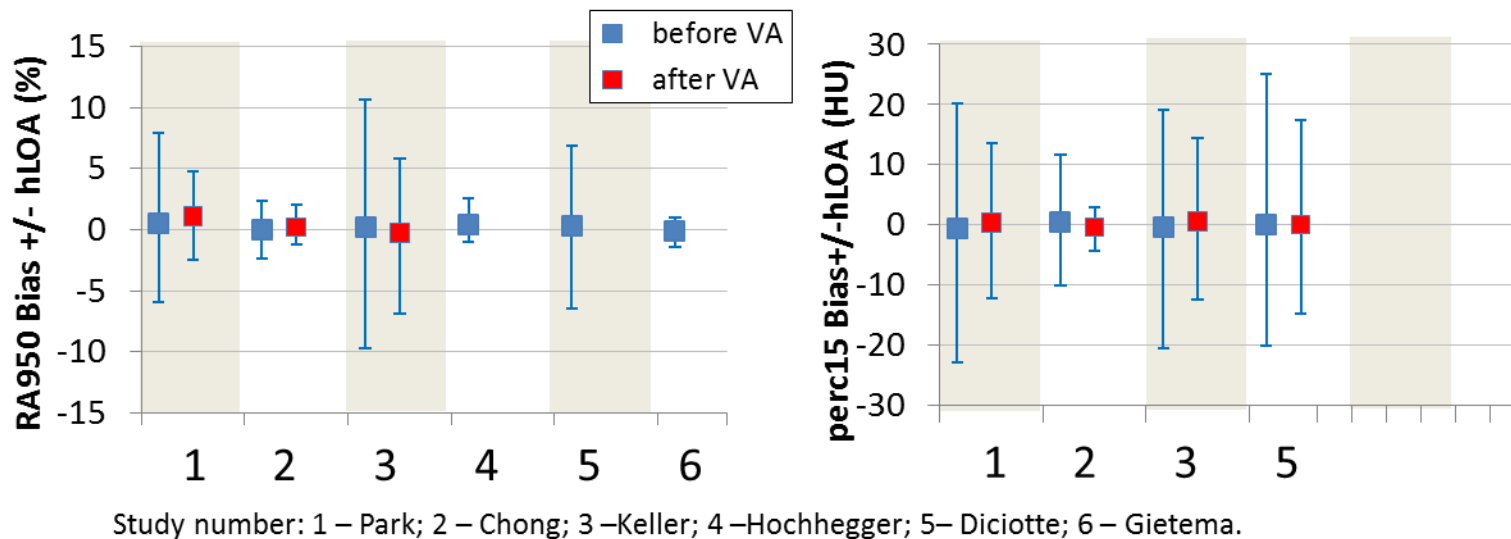


Fig. 1. Bias and limits of agreement compiled from the selected 6 studies: Bias and 95% limits of agreement (hLOA, or half-width of the 95% LOA, defined as $1.96 SD_{bias}$) for RA-950 HU (left panel) and Perc15 (right panel) both before (blue) and after (red) VA for selected studies (Table 1). The respective hLOAs are plotted as error bars. Study 3 reported RA -910 rather than RA-950. Studies 4, 5, and 6 did not perform VA for RA-950. Studies 4 and 6 did not report Perc15 results. Meta-analysis was not performed for RA-950 after VA because there are only two data points (excluding study 3), and one of which (study 2) reported statistically insignificant improvement of precision as a result of VA. For Perc15, meta-analysis was performed both before and after VA based on the 4 data points included.

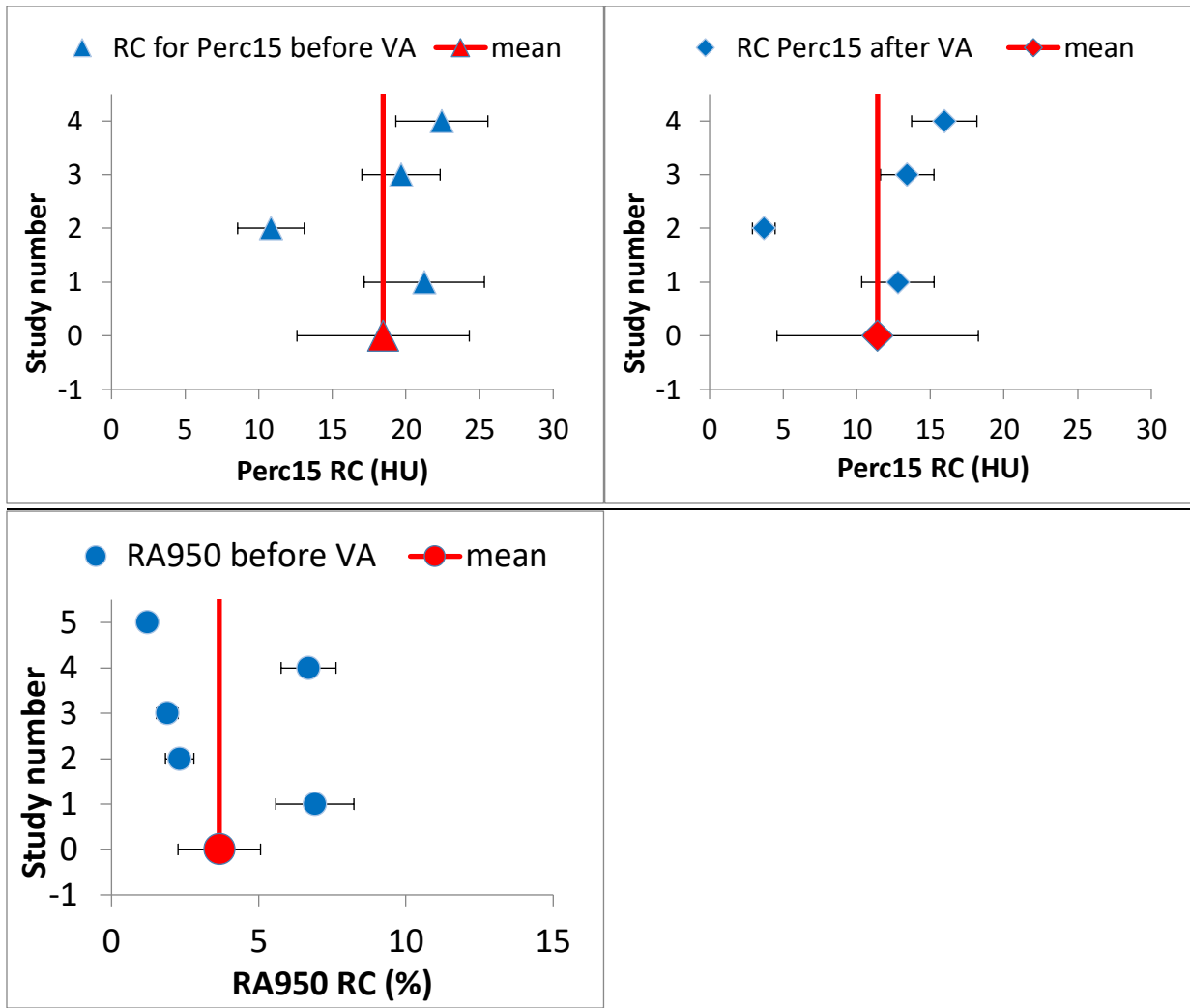


Fig 2. Forest plots of the repeatability coefficient (RC) of the studies (blue) and their pooled RC (red), with the 95% CI (defined as $1.96 \sqrt{SE_RC}$) shown as error bars: (a) Perc15 before VA, (b) Perc15 after VA, (c) RA-950 before VA. For RA-950, only two studies performed VA, and the effect of VA on the bias and precision were statistically insignificant.

995 **Appendix C: Conventions and Definitions**

Constructing the repeatability coefficient (RC) based on reported bias and limits of agreement

The studies reported either the standard deviation of the difference (SD_{bias}) or the limits of agreement (LOA) where the half width of LOA (hLOA) is defined as $1.96 SD_{bias}$. Conceptually this represents the variability remaining when the between- and within- subjects variability has been accounted for⁴⁰. For the assessment of repeat measurements on N subjects, we use the total within-subject standard deviation, wSD, defined below. The repeatability coefficient RC is defined as³⁸ $1.96 \sqrt{2} wSD$. Based on the definitions, the following relationship can be used to calculate the RC given the bias and hLOA:

$$wSD = \sqrt{\sum_i \frac{1}{2N} (Y_{i1} - Y_{i2})^2}$$

$$RC = 1.96\sqrt{2}wSD$$

$$hLOA = 1.96 \sqrt{\frac{1}{N-1} \sum_i (Y_{i1} - Y_{i2})^2 - \frac{N}{N-1} [(\overline{Y_1 - Y_2})]^2}$$

where

$$(\overline{Y_1 - Y_2}) = \bar{Y}_1 - \bar{Y}_2 = bias$$

and

$$\bar{Y}_1 = \sum_i^N \frac{Y_{i1}}{N}$$

$$\bar{Y}_2 = \sum_i^N \frac{Y_{i2}}{N}$$

are the means over N subjects for measurements 1 and 2, respectively.

Therefore,

$$RC^2 = \frac{N-1}{N} hLOA^2 + 1.96^2 bias^2$$

Where subscript $i = 1$ to N represents the subjects, and subscripts 1, 2 represent measurements at time points 1 and 2, respectively. The mean values are taken over the sample size N . When

1020 the bias is small, and N is large, RC approaches hLOA. For these studies, the difference is negligible, but the conceptual distinction is asserted here.

In addition, an uncertainty has been assigned to the RC itself for each study, based on the gamma distribution with shape and scale parameters defined by the sample size and the RC value itself³⁷. The variance of RC^2 is $Var(RC^2)=kb^2$, where $k =N(P-1)/2$ is the shape parameter (P
1025 is the number of independent measurements, or studies included in the meta-analysis), and $b=RC^2/k$ is the scale parameter. Using the “delta method”, one can show that $Var(RC)=Var(RC^2)/(4 RC^2)$.

Meta-analysis of the performance metric from the four studies

1030 The goal of the meta-analysis is to derive a mean RC based on the four studies selected. Following³⁷, the Perc15 or RA-950 metric data from the four studies were evaluated first using the fixed effects model assuming normal distribution as well as fixed effects model with exact maximum likelihoods. However, the heterogeneity test statistic I^2 approaches 100, requiring
1035 that the random effects model be used. A variance η^2 representing the underlying distribution of the RC's is constructed based on the DerSimonian and Laird estimator for this group of studies³⁷, which is added to the variance of each study to modify the weighting factor for the mean RC. The standard error for the mean RC thus obtained is also modified accordingly. This task was performed using Excel (Microsoft, Redmond WA). The same method was also used
1040 on the bias of these studies, which were homogeneously small and therefore a simple fixed effect model sufficed. VA does not have any statistically significant effect on the bias.

Appendix D: Model-specific Instructions and Parameters

1045 For acquisition modalities, reconstruction software and software analysis tools, profile conformance requires meeting the activity specifications above in Sections 2, 3 and 4. Sample acquisition and reconstruction protocols that are designed around the requirements of this Profile are provided for several common scanner makes and models in the Excel file “CT Protocols QIBA” linked at: [QIDW](#).

1050 Just using these parameters without meeting the requirements specified in the profile is not sufficient to achieve conformance. Conversely, it is possible to use different compatible parameters and still achieve conformance.

Sites using models listed here are encouraged to consider using these parameters for both simplicity and consistency. Sites using models not listed here may be able to devise their own settings that result in data meeting the requirements.

1055 **IMPORTANT: The presence of a product model/version in these tables does not imply it has demonstrated conformance with the QIBA Profile. Refer to the specifications and procedures in Sections 3 and 4, and the QIBA Conformance Statement for the product if available.**

1060 **Appendix E: Image Analysis Software Comparison**

This Appendix provides analysis software parameters that are expected to be compatible with the profile requirements.

1065 Research, open source and commercial vendors were invited to participant anonymously. The Radiological Society of North America (RSNA) acted as a neutral broker between vendors and the QIBA Lung Density committee to ensure the committee was blinded to vendor identify. All vendors indicated if their software was for research use/open-source or commercial. Vendors were instructed to generate measurements: 1) without segmentation quality assurance (QA) and manual correction to evaluate inter-software reproducibility; 2) repeated on the same
1070 images to evaluate intra-software reproducibility; and, 3) repeated on the same images with segmentation QA and manual correction.

The inter-vendor reproducibility coefficient (RDC)³⁹ was generated to measure the reproducibility across the different vendors for each measurement at standard dose: 1) with and without QA, and 2) by vendor type (research/open-source, commercial; Table 1). The RDC is the value under which the difference between repeated measurements on the same patient acquired under different conditions should fall within 95% probability, and is defined as:

$$RDC = 1.96\sqrt{2\sigma^2} \quad (1)$$

1075 where σ^2 is the mean of the variances of repeated measurements on the same patient. Low RDC values indicate high reproducibility between vendors. 95% confidence intervals for the RDC were constructed using bootstrapping with 5000 resamples. Table 1 shows the RDC for TLV, RA-950 and Perc15 by vendor type (research/open source or commercial) without and with QA. Research and commercial vendors RDC was comparable for TLV, RA-950 and Perc15 measurements: 0.39L / 0.32L, 1.2% / 1.2%, and 1.7HU / 1.6 HU, respectively. As shown in Table 1, QA had minimal impact on measurement reproducibility between vendors.
1080

1085 A total of 8 vendors participated in the software comparison study; n=4 research/open-source and n=4 commercial. A single commercial vendor withdrew from the study. To estimate the RDC for a given vendor we must estimate the variance relative to the other K-1 vendors in the comparison (K = 8 in our study). Therefore, for specific vendor, l , we calculate the mean variance, σ_l^2 , for the measurements, subscript i , across the 50 image sets. Let $M_{i,k}$ represent measurement i of vendor l .
1090

Let $\sigma_{i,k,l}^2$ represent the variance between vendor l and vendor k for measurement i .

$$\sigma_{i,k,l}^2 = \frac{1}{2}(M_{i,k} - M_{i,l})^2$$

1095 Let $\sigma_{k,l}^2$ represent the variance between vendor k and vendor l averaged over all measurements N .

$$\sigma_{k,l}^2 = \frac{1}{N} \sum_{i=1}^N \sigma_{i,k,l}^2 = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} (M_{i,k} - M_{i,l})^2 = \frac{1}{2N} \sum_{i=1}^N (M_{i,k} - M_{i,l})^2$$

Now average the variance over the other $K - 1$ vendors to generate the average variance for

vendor l .

1100

$$\sigma_l^2 = \frac{1}{K-1} \sum_{k=1}^{K-1} \sigma_{k,l}^2$$

The average RDC for vendor l is then given by:

$$RDC_l = 1.96 * \sqrt{2\sigma_l^2}$$

1105

Table 2 shows the RDC_l for TLV, RA-950 and Perc15 for 8 different software vendors with and without quality assurance (QA) using manual correction of the lung volume segmentation. Overall, inter-software RDC was low at 0.35L, 1.2% and 1.8HU for TLV, RA-950 and Perc15, respectively. For all vendors, inter-software RDC remained unchanged following QA: 0.35L, 1.2% and 1.8HU for TLV, RA-950 and Perc15, respectively.

1110

Intra-software RDC was also generated by having the vendors perform repeated measurements without QA; all vendors had an intra-software RDC of 0 indicating the repeated application of the identical software algorithm is deterministic as expected.

1115

Table 1. The RDC for TLV, LAA950 and Perc15 for Research/Open-source and Commercial Vendors for Standard Dose

Parameter	Inter-software RDC without QA	95% CI	Inter-software RDC with QA	95% CI
TLV (L)				
Research/Open Source	0.39	0.36 – 0.41	0.39	0.36 – 0.41
Commercial	0.32	0.29 – 0.34	0.32	0.29 – 0.35
LAA950 (%)				
Research/Open Source	1.2	0.9 – 1.4	1.2	0.9 – 1.4
Commercial	1.2	1.0 – 1.3	1.1	1.0 – 1.3
Perc15 (HU)				
Research/Open Source	1.7	1.5 – 1.9	1.7	1.5 – 1.9
Commercial	1.6	1.3 – 1.9	1.6	1.3 – 2.0

Table 2. The RDC for TLV, LAA950 and Perc15 for All Vendors at Standard Dose with and without QA

Parameter	Inter-software RDC Without QA		Inter-software RDC With QA	
	RDC	95% CI	RDC	95% CI
TLV (L)				
Total	0.35	0.32 – 0.37	0.35	0.32 – 0.37
Vendor 1	0.38	0.35 – 0.41	0.38	0.35 – 0.42
Vendor 2	0.26	0.24 – 0.27	0.26	0.24 – 0.28
Vendor 3	0.26	0.24 – 0.29	0.26	0.24 – 0.29
Vendor 4	0.48	0.46 – 0.51	0.48	0.45 – 0.51
Vendor 5	0.25	0.23 – 0.27	0.25	0.23 – 0.27
Vendor 6	0.46	0.43 – 0.49	0.46	0.43 – 0.49
Vendor 7	0.31	0.28 – 0.34	0.31	0.28 – 0.34
Vendor 9	0.26	0.24 – 0.28	0.26	0.24 – 0.28
LAA950 (%)				
Total	1.2	1.0 – 1.4	1.2	1.0 – 1.4
Vendor 1	1.2	1.0 – 1.5	1.2	1.0 – 1.5
Vendor 2	1.1	0.9 – 1.2	1.1	0.9 – 1.2
Vendor 3	1.1	0.9 – 1.2	1.1	0.9 – 1.2
Vendor 4	1.2	0.9 – 1.4	1.2	0.9 – 1.4
Vendor 5	1.2	1.0 – 1.3	1.2	1.0 – 1.3
Vendor 6	1.5	1.2 – 1.8	1.5	1.2 – 1.8
Vendor 7	0.9	0.7 – 1.0	0.9	0.7 – 1.0
Vendor 9	1.2	1.0 – 1.4	1.2	1.0 – 1.4
Perc15 (HU)				
Total	1.8	1.6 – 2.0	1.8	1.6 – 2.1
Vendor 1	1.6	1.4 – 1.9	1.7	1.4 – 1.9
Vendor 2	1.5	1.3 – 1.7	1.6	1.3 – 1.8
Vendor 3	1.5	1.3 – 1.6	1.5	1.3 – 1.6
Vendor 4	2.3	2.1 – 2.6	2.3	2.1 – 2.6
Vendor 5	2.1	1.9 – 2.3	2.1	1.9 – 2.3
Vendor 6	2.0	1.6 – 2.3	2.0	1.6 – 2.4
Vendor 7	1.4	1.2 – 1.7	1.4	1.2 – 1.6
Vendor 9	1.7	1.5 – 1.9	1.7	1.5 – 1.9

Appendix F. Assessment procedure for computing reproducibility of Lung Density Software across various vendors

1125 Inter-software reproducibility refers to the ability of different Lung Density software packages to reproduce Lung Density metrics. Variations in Lung Density metrics such as RA-950 and Perc15 result from differences in the underlying lung segmentation algorithms and/or post-processing techniques used by different software packages.

1130 Inter-software reproducibility of Lung Density is measured using the reproducibility coefficient (RDC)³⁹. The reproducibility coefficient is defined in metrology guidelines as “the least significant difference between two repeated measurements taken under different conditions.”⁴⁵. In this case, the “different conditions” are the different software packages.

1135 To quantify the RDC associated with a new software package for measuring Lung Density, a reference data set may be analyzed with the new software and results compared to performance of established commercial and open source software packages. The reference data set is made available for download upon request at the quantitative imaging data warehouse (QIDW). The reference data set consists of 50 subjects with varying degrees of airflow limitation, ranging GOLD 0 (n=10) to GOLD 4 COPD (n=10 cases in each GOLD group).

1140 There are two scans for each subject, one with conventional x-ray dose (~6 mGy average CTDIvol) and one with reduced dose x-ray (~3 mGy average CTDIvol). Total lung volume, and RA-950 and Perc15 were previously measured by 8 commercial software vendors and open source academic centers. Although no perfect ground truth for these measurements was available, a consensus measurement for each scan may be generated from the RDC values for lung volume, RA-950 and Perc15 from all vendors, which are included in Appendix E.

1150 To qualify a new vendor software algorithm (“Vendor A”), the performance for the measurements, M_i , using Vendor A’s software algorithm must be compared to the consensus mean measurement, \bar{M}_i , for the reference data set described in the previous paragraph, where the subscript i refers to the index for the i th image set from the 50 subjects.

The RDC is defined according to section 7.2 of Obuchowski et al⁴⁵ as:

$$RDC = 1.96 * \sqrt{2\sigma_{inter-software}^2}$$
$$\sigma_{inter-software}^2 = \frac{1}{N-1} \sum_{i=1}^N (M_i - \bar{M}_i)^2$$
$$\bar{M}_i = \frac{1}{K} \sum_{k=1}^K M_k$$

1155 where N is the total number of scans (50 in our example), K is the total number of software algorithms represented in the reference data set (8 in our study), and M_{ik} is the lung density (or lung volume) measurement for i th image set and k th algorithm.

1160 Note that 50 full-dose and 50 low-dose scans exist for the purposes of computing the RDC of a new software package with multiple dose levels. RDC should be computed separately for both dose levels.

1165 Alternatively, and for the purposes of research studies or clinical trials, it may be of interest to compute the RDC between two specific software packages rather than the RDC of a new software package compared against the vendor consensus. One scenario where this may occur is for a longitudinal study where one software package was used to compute Lung Density measures for baseline scans but a different software package was used for follow-up scans. In that case, data from the QIDC warehouse or study specific data should be processed using both
1170 software packages and the RDC computed using the above equation.

In practice, longitudinal studies that utilize multiple software packages will need to consider both the within-subject variability (repeatability) and the inter-software reproducibility. The overall reproducibility, which considers both factors, can be estimated as:

1175

$$RDC_{overall} = 1.96 * \sqrt{2(wSD^2 + \sigma_{inter-software}^2)}$$

Where wSD^2 is within-subject variance explained in Appendix C and σ^2 is the inter-software variance.

References for Appendices A-F:

- 1180 1. Castaldi PJ, San Jose Estepar R, Mendoza CS, et al. Distinct quantitative computed tomography emphysema patterns are associated with physiology and function in smokers. *Am J Respir Crit Care Med*. 2013;188(9):1083-1090.
2. Schroeder JD, McKenzie AS, Zach JA, et al. Relationships between airflow obstruction and quantitative CT measurements of emphysema, air trapping, and airways in subjects with and without chronic obstructive pulmonary disease. *AJR Am J Roentgenol*. 2013;201(3):W460-470.
- 1185 3. Van Tho N, Wada H, Ogawa E, Nakano Y. Recent findings in chronic obstructive pulmonary disease by using quantitative computed tomography. *Respir Investig*. 2012;50(3):78-87.
- 1190 4. Stolk J, Putter H, Bakker EM, et al. Progression parameters for emphysema: a clinical investigation. *Respir Med*. 2007;101(9):1924-1930.
5. Martinez CH, Chen YH, Westgate PM, et al. Relationship between quantitative CT metrics and health status and BODE in chronic obstructive pulmonary disease. *Thorax*. 2012;67(5):399-406.
- 1195 6. Pare PD, Camp PG. Airway disease and emphysema on CT: not just phenotypes of lung pathology. *Thorax*. 2012;67(5):380-382.
7. Hayhurst MD, MacNee W, Flenley DC, et al. Diagnosis of pulmonary emphysema by computerised tomography. *Lancet*. 1984;2(8398):320-322.
- 1200 8. Bankier AA, De Maertelaer V, Keyzer C, Gevenois PA. Pulmonary emphysema: subjective visual grading versus objective quantification with macroscopic morphometry and thin-section CT densitometry. *Radiology*. 1999;211(3):851-858.
9. Gevenois PA, De Vuyst P, de Maertelaer V, et al. Comparison of computed density and microscopic morphometry in pulmonary emphysema. *Am J Respir Crit Care Med*. 1996;154(1):187-192.
- 1205 10. Yuan R, Hogg JC, Pare PD, et al. Prediction of the rate of decline in FEV(1) in smokers using quantitative Computed Tomography. *Thorax*. 2009;64(11):944-949.
11. Moyer VA, Force USPST. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;160(5):330-338.
12. Madani A, Zanen J, de Maertelaer V, Gevenois PA. Pulmonary emphysema: objective quantification at multi-detector row CT--comparison with macroscopic and microscopic morphometry. *Radiology*. 2006;238(3):1036-1043.
- 1210 13. Marsh S, Aldington S, Williams MV, et al. Physiological associations of computerized tomography lung density: a factor analysis. *Int J Chron Obstruct Pulmon Dis*. 2006;1(2):181-187.
- 1215 14. Bakker ME, Putter H, Stolk J, et al. Assessment of regional progression of pulmonary emphysema with CT densitometry. *Chest*. 2008;134(5):931-937.
15. Chong D, Brown MS, Kim HJ, et al. Reproducibility of volume and densitometric measures of emphysema on repeat computed tomography with an interval of 1 week. *Eur Radiol*. 2012;22(2):287-294.
- 1220 16. Hochegger B, Irion KL, Marchiori E, Moreira JS. Reconstruction algorithms and their influence in emphysema CT measurements. *Acad Radiol*. 2010;17(5):674.

17. Hochegger B, Irion KL, Marchiori E, Moreira JS. Reconstruction algorithms influence the follow-up variability in the longitudinal CT emphysema index measurements. *Korean J Radiol.* 2011;12(2):169-175.
- 1225 18. Madani A, Van Muylem A, Gevenois PA. Pulmonary emphysema: effect of lung volume on objective quantification at thin-section CT. *Radiology.* 2010;257(1):260-268.
19. Dirksen A. Monitoring the progress of emphysema by repeat computed tomography scans with focus on noise reduction. *Proc Am Thorac Soc.* 2008;5(9):925-928.
- 1230 20. Keller BM, Reeves AP, Henschke CI, Yankelevitz DF. Multivariate compensation of quantitative pulmonary emphysema metric variation from low-dose, whole-lung CT scans. *AJR Am J Roentgenol.* 2011;197(3):W495-502.
21. Bakker ME, Stolk J, Putter H, et al. Variability in densitometric assessment of pulmonary emphysema with computed tomography. *Invest Radiol.* 2005;40(12):777-783.
- 1235 22. Park SJ, Lee CH, Goo JM, Heo CY, Kim JH. Inter-scan repeatability of CT-based lung densitometry in the surveillance of emphysema in a lung cancer screening setting. *Eur J Radiol.* 2012;81(4):e554-560.
23. Stoel BC, Putter H, Bakker ME, et al. Volume correction in computed tomography densitometry for follow-up studies on pulmonary emphysema. *Proc Am Thorac Soc.* 2008;5(9):919-924.
- 1240 24. Stockley RA, Parr DG, Piitulainen E, Stolk J, Stoel BC, Dirksen A. Therapeutic efficacy of alpha-1 antitrypsin augmentation therapy on the loss of lung tissue: an integrated analysis of 2 randomised clinical trials using computed tomography densitometry. *Respiratory research.* 2010;11:136.
- 1245 25. Coxson HO, Dirksen A, Edwards LD, et al. The presence and progression of emphysema in COPD as determined by CT scanning and biomarker expression: a prospective analysis from the ECLIPSE study. *Lancet Respir Med.* 2013;1(2):129-136.
26. Gietema HA, Schilham AM, van Ginneken B, van Klaveren RJ, Lammers JW, Prokop M. Monitoring of smoking-induced emphysema with CT in a lung cancer screening setting: detection of real increase in extent of emphysema. *Radiology.* 2007;244(3):890-897.
- 1250 27. Soejima K, Yamaguchi K, Kohda E, et al. Longitudinal follow-up study of smoking-induced lung density changes by high-resolution computed tomography. *Am J Respir Crit Care Med.* 2000;161(4 Pt 1):1264-1273.
28. Stolk J, Dirksen A, van der Lugt AA, et al. Repeatability of lung density measurements with low-dose computed tomography in subjects with alpha-1-antitrypsin deficiency-associated emphysema. *Invest Radiol.* 2001;36(11):648-651.
- 1255 29. Shaker SB, Dirksen A, Laursen LC, et al. Short-term reproducibility of computed tomography-based lung density measurements in alpha-1 antitrypsin deficiency and smokers with emphysema. *Acta Radiol.* 2004;45(4):424-430.
30. McGregor A, Roberts HC, Dong Z, et al. Repeated low-dose computed tomography in current and former smokers for quantification of emphysema. *J Comput Assist Tomogr.* 2010;34(6):933-938.
- 1260 31. Parr DG, Sevenoaks M, Deng C, Stoel BC, Stockley RA. Detection of emphysema progression in alpha 1-antitrypsin deficiency using CT densitometry; methodological advances. *Respir Res.* 2008;9:21.

- 1265 32. Hoffman EA, Jiang R, Baumhauer H, et al. Reproducibility and validity of lung density measures from cardiac CT Scans--The Multi-Ethnic Study of Atherosclerosis (MESA) Lung Study. *Acad Radiol*. 2009;16(6):689-699.
33. Mets OM, Isgum I, Mol CP, et al. Variation in quantitative CT air trapping in heavy smokers on repeat CT examinations. *Eur Radiol*. 2012;22(12):2710-2717.
- 1270 34. Stoel BC, Bakker ME, Stolk J, et al. Comparison of the sensitivities of 5 different computed tomography scanners for the assessment of the progression of pulmonary emphysema: a phantom study. *Invest Radiol*. 2004;39(1):1-7.
35. Stoel BC. Personal Communication. 2014.
36. Staring M, Bakker ME, Stolk J, Shamonin DP, Reiber JH, Stoel BC. Towards local progression estimation of pulmonary emphysema using CT. *Med Phys*. 2014;41(2):021905.
- 1275 37. Winter B. *Linear models and linear mixed effects models in R with linguistic applications*.: arXiv:1308.5499;2013.
38. Huang EP, Wang XF, Choudhury KR, et al. Meta-analysis of the technical performance of an imaging procedure: guidelines and statistical methodology. *Stat Methods Med Res*. 2015;24(1):141-174.
- 1280 39. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. *Stat Methods Med Res*. 2015;24(1):27-67.
- 1285 40. Kessler LG, Barnhart HX, Buckler AJ, et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. *Stat Methods Med Res*. 2015;24(1):9-26.
41. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. *Am J Phys Med Rehabil*. 2005;84(9):719-723.
- 1290 42. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231-240.
43. Tesio L. Outcome measurement in behavioural sciences: a view on how to shift attention from means to individuals and why. *Int J Rehabil Res*. 2012;35(1):1-12.
44. Diciotti S, Sverzellati N, Kauczor HU, et al. Defining the intra-subject variability of whole-lung CT densitometry in two lung cancer screening trials. *Acad Radiol*. 2011;18(11):1403-1411.
- 1295 45. Obuchowski NA, Reeves AP, Huang EP, et al. Quantitative imaging biomarkers: a review of statistical methods for computer algorithm comparisons. *Stat Methods Med Res*. 2015;24(1):68-106.

1300