QIBA Profile:
Computed Tomography: Lung Densitometry

Table of Contents

Open Issues: 1
Closed Issues: 3
1. Executive Summary 5
2. Clinical Context and Claims 6
3. Profile Activities 9
3.1. Product Qualification 11
3.1.1 Discussion 11
3.1.2 Specification 12
3.2. Staff Qualification 14
3.3. Periodic QA 14
3.3.1 Discussion 14
3.3.2 Specification 14
3.4. Protocol Design 14
3.5. Subject Handling 18
3.5.1 Discussion 18
3.5.2 Specification 19
3.6. Image Data Acquisition 20
3.6.1 Discussion 20
3.6.2 Specification 20
3.7. Image Data Reconstruction 21
3.7.1 Discussion 21
3.7.2 Specification 21
3.8. Image QA 22
3.8.1 Discussion 22
3.8.2 Specification 22
3.9. Image Analysis 23
3.9.1 Discussion 23
3.9.2 Specification 24
3.10. Image Interpretation 25
3.10.1 Discussion 25
4. Assessment Procedures 25
4.1.1 Assessment Procedure: HU Bias and Repeatability 27
4.1.2 Assessment Procedure: Voxel Noise 28
4.1.3 Assessment Procedure: In-Plane Spatial Resolution and Edge Enhancement 29
5. Conformance 30
References 32
Appendices 34
Appendix A: Acknowledgements and Attributions 34
Appendix B: Conformance Checklists 34
Acquisition Device Checklist 34
Image Analysis Software Checklist 35
Radiologist Checklist 36
Physicist Checklist 37

Technologist Checklist 38
Clinician OR Statastician Checklist 40
Appendix C: Background Information 41
Appendix D: Conventions and Definitions 56
Appendix E: Model-specific Instructions and Parameters 58
Appendix F: Image Analysis Software Comparison 59
Appendix G. Assessment procedure for computing reproducibility of Lung Density Software across various vendors 62
References for Appendices A-G: 64

Open Issues:

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.

1. "Iterative Reconstruction"

Q. What is the effect of iterative reconstruction (IR) methods on RA-950 HU and Perc15 measures of lung parenchymal density?

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.

2. "Harmonization"

Q. What is the best reference standard for harmonizing systematic differences in quantitative CT number (Hounsfield Unit – HU) across scanner make and model?

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 4.1.2) 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 link to protocol examples in Appendix E). 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 ± 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. Standardizing CTDIvol using water equivalent diameter would be superior if readily feasible. However, at present such an approach would require additional off-line calculations that are unlikely to be robustly integrated into a multi-center trial at the present time.

Closed Issues:

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. "Breath-hold Consistency"
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?

A. Yes. Based on review of Dr. Park's study in the meta-analysis, the range of fractional change of volume, V2/V1, 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.

5. "Cross-Sectional Claim"
Q. Will there be a cross-sectional claim?

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.

6. "**Specification of matrix size**"

Q. Should the acquisition parameters include specification of matrix size or display FOV?

A. This is specified if the in-plane spatial resolution is met.

1. Executive Summary

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 C 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.
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.

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 Product Qualification, Staff Qualification, Periodic Quality Assurance, Subject Handling, Protocol Design, Image Data Acquisition, Image Data Reconstruction, Image QA, Image Distribution and 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 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 ≥ 3.7% of the normalized lung volume, or a change in Perc15 of ≥ 11 HU after lung volume adjustment can be accepted as indicative of a true change (with 95% confidence).

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 patient care.

The compilation of this document represents the efforts of many individuals over a several years of effort, some but not all of whom are acknowledged in Appendix A. QIBA Profiles addressing other imaging biomarkers using CT, MRI, PET and Ultrasound can be found at qibawiki.rsna.org.

2. Clinical Context and Claims

Clinical Context

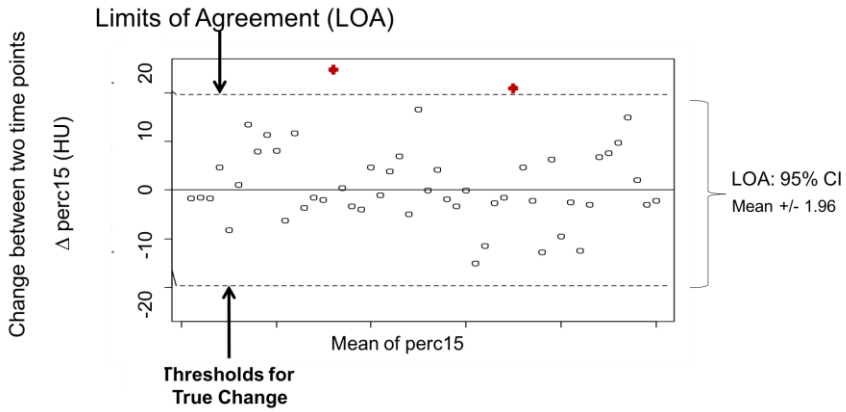

Figure 1: Example Bland-Altman plot for perc15 without VA; the LOA (or RC when there's no bias) are the thresholds required for detection of a true increase in the extent of emphysema. The first step is to assess whether a measurement meets the threshold for a true change (red crosses above the upper 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 (Perc15).

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

Claim 1: With lung volume adjustment (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.

Claim 2: Without VA, a decrease in Perc15 of at least 18 HU indicates an increase in the extent of emphysema, with 95% confidence.

Claim 3: Without VA, an increase in RA-950* of at least 3.7% indicates 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 [Δx -18 HU, Δx +18 HU] with 95% confidence; and for a measured change of Δy% in RA-950, the true change is expected to lie in the interval [Δy -3.7%, Δy +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 [Δx -11 HU, Δx +11 HU] with 95% confidence. 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 √2 wSD, where wSD is the within-subject standard deviation. 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 of studies conducted at the same site using the same scan protocol and CT scanner make and model described in Appendix C. 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. 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. 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 a link to specific protocols in Appendix E that if combined with the requirements on the other actors, will meet or exceed the claims. More details on how limits of agreement are calculated based on a repeatability meta-analysis, including a description of VA methods that fit the selection criteria are provided below and in Appendix C.

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 ≥ 3.7% of the normalized lung volume, or a change in Perc15 of ≥ 11 HU after lung volume adjustment 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 HU value is readily translated into a more general measure of density, grams per liter (g/L) using a simple formula:

$$Density\ (grams\ per\ liter) = Attenuation\ (Hounsfield\ Units) + 1000.$$

Because lung volume is a critical determinant of lung density, the conversion from Hounsfield units to grams per liter is usually accompanied by adjustment for lung volume (see also Appendix C). More specifically:

$$Volume\ adjusted\ lung\ density\ Perc15 = Measured\ lung\ density\ Perc15 \cdot \left(\frac{Measured\ CT\ Lung\ Volume}{Predicted\ CT\ Lung\ Volume}\right)$$

where the measured Perc15 is the value measured from the CT attenuation histogram, the measured CT lung volume is the number of voxels in the lungs after segmentation from the surrounding structures multiplied by the voxel volume, and the predicted CT lung volume is that predicted from equations generated from normal healthy individuals with no smoking history.

The committee recognizes that the limits of variability reported in the claims 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.

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	Product Qualification	3.1
	Periodic QA	3.3
	Subject Handling	3.5
	Image Data Acquisition	3.6
Radiologist	Staff Qualification	3.2
	Protocol Design	3.4
	Subject Handling	3.5
Physicist	Product Qualification	3.1

	Staff Qualification	3.2
	Periodic QA	3.3
	Protocol Design	3.4
	Image QA	3.8
Technologist	Subject Handling	3.5
	Image Data Acquisition	3.6
	Image Data Reconstruction	3.7
	Image QA	3.8
Clinician	Image Analysis	3.9
Reconstruction Software	Image Data Reconstruction	3.7
Image Analysis Software	Image Analysis	3.9
Statistician	Image Interpretation	3.10

The requirements in this Profile do not codify a Standard of Care; they only provide guidance intended to achieve the stated Claim. Failing to conform to a "shall" in this Profile is a protocol deviation. Although deviations invalidate the Profile Claim, such deviations may be reasonable, and unavoidable, and the radiologist or supervising physician is expected to do so when required by the best interest of the patient or research subject. How 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**.

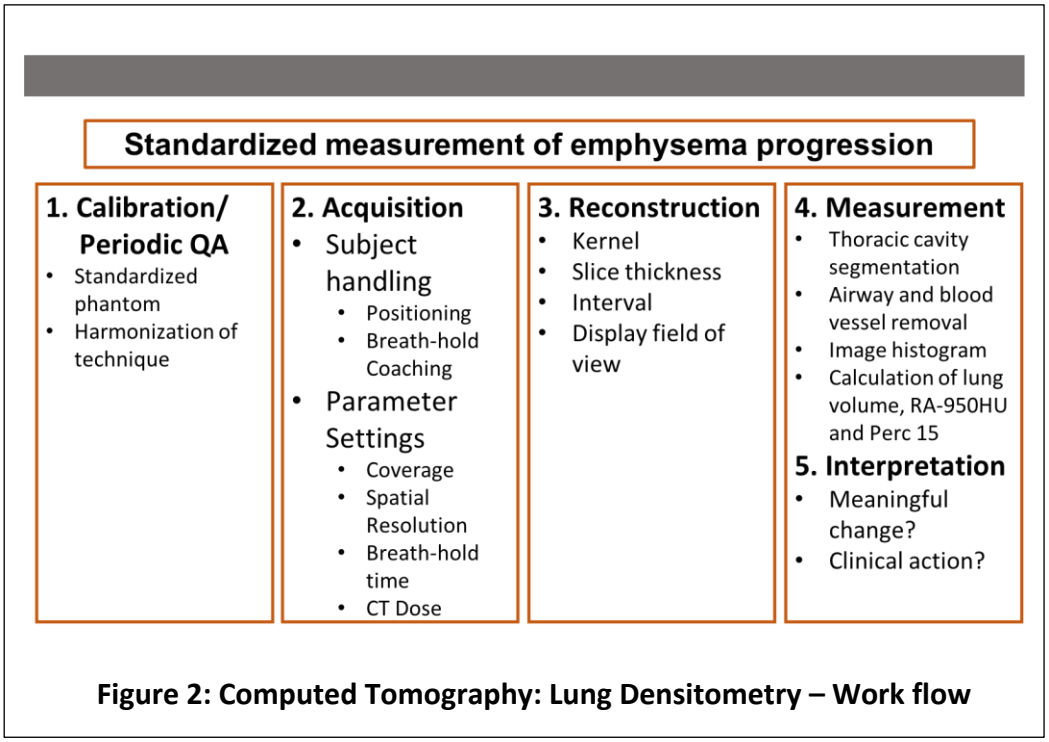

Figure 2: Computed Tomography: Lung Densitometry – Work flow

3.1. Product Qualification

This activity describes specifications for performance assessment, calibration or standardization, and validations of equipment that are necessary to reliably meet the Profile Claim. Procedures for meeting these specifications are as described in Sections 4.1.1-4.1.3.

3.1.1 DISCUSSION

These specifications are defined based on groundwork projects from vendor round 1 and 2 studies (See "Section 4: Assessment Procedures**"**) and the QIBA-SRM phantoms[6]. 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 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. Parameter value setting (i.e. table speed, rotation time, collimation, pitch) must cover an axial field of view of 35 cm in 10 seconds or less.

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.

A model-based correction method based on the composition, and the certified physical densities of the constituent materials, of the QIBA-SRM phantom[6] 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 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 longitudinal change in lung density requires calibration of Hounsfield Unit (HU) values to improve precision and reduce bias. 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 as described in the assessment procedure in Section 4.1.1 using the COPD Gene, or similarly designed, phantom where uniform low density regions are provided. 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 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 **Appendix F**.

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 ≤ ± 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, Assessment Procedure: In-Plane, Through-Plane (Z-axis) Spatial Resolution and Edge Enhancement.
Through-plane spatial resolution	Acquisition Device	Shall demonstrate a slice sensitivity profile with FWHM ≤ 1.0 mm as described in Section 4.1.3, Assessment Procedure: In-Plane, Through-Plane (Z-axis) Spatial Resolution and Edge Enhancement .
Edge Enhancement	Acquisition Device	Shall demonstrate an edge enhancement ≤ 3% for the edge response function as described in Section 4.1.3, Assessment Procedure: In-Plane, Through-Plane (Z-axis) Spatial Resolution and Edge Enhancement.
Acquisition speed	Acquisition Device	Shall set parameter values that will cover an axial field of view of 35 cm in 10 seconds or less.
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) as described in Section 4.1.1, Assessment Procedure: HU Bias and Repeatability.
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 as described in Section 4.1.1, Assessment Procedure: HU Bias and Repeatability.
Lung Density Analysis	Image Analysis Software	Shall calculate and output for the whole lung: ● RA-950 HU ● Perc15 ● Lung Density Histogram ● Total Lung Volume As described in Section 4.2, Assessment Procedure: Reproducibility of Image Analysis Software across Various Vendors.
Reproducibility of Analysis Software	Image Analysis Software	Shall use identical measurement algorithm for each longitudinal time point measured. See Section 4.2, Assessment Procedure:

		Reproducibility of Image Analysis Software across Various Vendors for more information.

3.2. Staff Qualification

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

3.2.1 DISCUSSION

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

3.3. Periodic QA

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

3.3.1 DISCUSSION

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

3.3.2 SPECIFICATION

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	Shall assess the CT conformance for the measured HU value and standard deviation in 3.1.2 are met using procedures in 4.1.1 on a monthly basis.
HU Stability	Physicist	Shall meet the specifications in Table 3.1.2.

3.4. Protocol Design

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

3.4.1 DISCUSSION

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

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.

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. This is intended to allow as much flexibility as possible for product innovation and reasonable adjustments for patient size (such as increasing acquisition mAs and reconstruction DFOV for larger patients), while reaching the performance targets. Again, the technique parameter sets in the Conformance Statements for Acquisition Devices and Reconstruction Software may be helpful for those looking for more guidance.

The purpose of the minimum **scan duration** requirement is to permit acquisition of 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.

Total Collimation Width (defined as the total nominal beam width, N x T, for example 64 detectors x 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.

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.

Smaller voxels are preferable to reduce partial volume effects and provide higher accuracy due to higher spatial resolution. The resolution/voxel size that reaches the analysis software is affected by both acquisition parameters and reconstruction parameters.

X-ray CT uses ionizing radiation. Exposure to radiation can pose risks; however, as the radiation dose is reduced, image quality can be degraded. It is expected that health care professionals will balance the need for good image quality with the risks of radiation exposure on a case-by-case basis. It is not within the scope of this document to describe how these trade-offs should be resolved, 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 ± 18% for subject weight between 50-100 kg [11], 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).

Image reconstruction is modeled as a separate Activity in the QIBA Profile. Although it is closely related to image acquisition, and is usually performed on the Acquisition Device, reconstruction may be performed, or re-performed, separate from the acquisition. Many reconstruction parameters will be influenced or constrained by related acquisition parameters. This specification is the result of discussions to allow a degree of separation in their consideration without suggesting they are totally independent.

Many reconstruction parameters can have direct or indirect effects on 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.

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 4.1.3. 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 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 Table 3.4.2. Examples of kernels that would meet the specifications for the major vendors are described in Section 4.1.3.

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

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 responsible for performing the validation. The role of the Physicist actor may be played by an in-house diagnostic radiology QMP, 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. Note that parameter value setting (i.e. table speed, rotation time, collimation, pitch) must cover an axial field of view of 35 cm in 10 seconds or less. Some parameters are system dependent and may require special attention from the Physicist actor or equivalent.

Parameter	Actor	Specification	DICOM Tag
Acquisition Protocol	Radiologist	Shall prepare a protocol to meet the specifications in this table.	
Acquisition Protocol	Radiologist	Shall ensure technologists have been trained on the requirements of this profile.	
Total Collimation Width	Radiologist	Shall set to Greater than or equal to 16 mm.	Total Collimation Width (0018,9307)
Nominal Tomographic Section Thickness	Radiologist	Shall set to Less than or equal to 1.0 mm using procedures in 4.1.3	Single Collimation Width (0018,9306)
Scan Duration	Radiologist	Shall set parameter values that will cover an axial field of view of 35 cm in 10 seconds or less.	Table Speed (0018,9309)
Reconstruction Protocol	Radiologist	Shall prepare a protocol to meet the specifications in this table.	
Reconstruction Protocol	Radiologist	Shall ensure technologists have been trained on the requirements of this profile.	
Reconstructed Image Thickness	Physicist	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 ≤ 1.0 mm using procedures in 4.1.3.	
Through-plane spatial resolution	Physicist	Shall validate that the protocol achieves a slice sensitivity profile with FWHM ≤ 1.0 mm using procedures in 4.1.3.	
Edge Enhancement	Physicist	Shall validate that the protocol achieves a minimum edge enhancement of 3% for the	

		edge response function as described in Section 4.1.3.	
Voxel Noise	Physicist	Shall validate that the protocol achieves a standard deviation of voxel noise that is ≤ 20HU for lung equivalent foam, air and water materials inside a phantom as described in Section 4.1.2.	

3.5. Subject Handling

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

3.5.1 DISCUSSION

Use of Contrast

Intravenous or oral contrast will interfere with the quantitative density measures and will prevent meeting the profile claims. It is the responsibility of the Radiologist to insure that contrast is not prescribed.

Subject Positioning

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

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.

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).

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, 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 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.

To extract the desired information from the CT images, it is very important that the breathing instructions are followed closely. Refer to this publication[1] for further visual and description information on proper breathing instructions.

An example of a breath-hold coaching script is:

- "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"
- "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."

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).
Scan Projection	Technologist	Shall perform a lateral scout and verify that the mid-

Radiograph		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.
Breath-hold Coaching	Technologist	Shall coach the subject on Breath-holding as described above.
Use of intravenous contrast	Radiologist	Shall NOT use intravenous contrast.
Use of oral contrast	Radiologist	Shall NOT use oral contrast.
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

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

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.

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").	
	Technologist	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 make and model with equivalent acquisition and reconstruction parameters to that of the baseline CT scan.	
Scan Projection Radiograph	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.	<Confirm DICOM Field> Reconstruction Field of View (0018,9317)
Axial field of view	Technologist	Shall match the display field of view to that of the Baseline scan, if available.	Reconstruction Diameter (0018,1100)

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

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 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.

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.

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.
- 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:

- 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.

3.8.2 SPECIFICATION

Parameter	Actor	Requirement
Adequate coverage	Technologist	Shall confirm the lung volume is fully represented in the field of 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
Conformance to baseline.	Technologist	Shall confirm the Protocol is consistent with the baseline.
Absence of IV	Technologist	Shall confirm the absence of IV contrast

contrast		
Lung Volume	Technologist	Shall confirm that the proper breath-hold coaching procedure outlined in section 3.5.2 is followed.
Image Distribution	Technologist	Images shall be transmitted in uncompressed DICOM format and according to the anonymization standards approved for the study.

If the image acquisition and software analysis do not conform to the above specifications, the profile performance specifications may no longer be valid.

3.9. 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.

3.9.1 DISCUSSION

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), manual segmentation is labor intensive necessitating automated segmentation techniques. 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 colleagues [13]. Other methods, such as statistical shape modeling and atlas-based segmentation have been proposed [14].

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 F.

3.9.2 SPECIFICATION

Parameter	Actor	Requirement
Lung Density Analysis	Image Analysis Software	Shall use a consistent lung segmentation procedure including the following steps. 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 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, Assessment Procedure: Reproducibility of Image Analysis Software across Various Vendors for more detail on procedures.
	Clinician	Shall calculate and output the whole lung RA-950, Perc15 and Lung Density Histogram: The analysis software used shall be identical for each longitudinal time point measured (reanalyze images if necessary), 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, Assessment Procedure: Reproducibility of Image Analysis Software across Various Vendors for more detail on procedures

3.10. 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.

3.10.1 DISCUSSION

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/year [3], and the mean change in cigarette smokers with COPD is about 1.1 g/L/year [5]. 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 emphysema [15].

4. Assessment Procedures

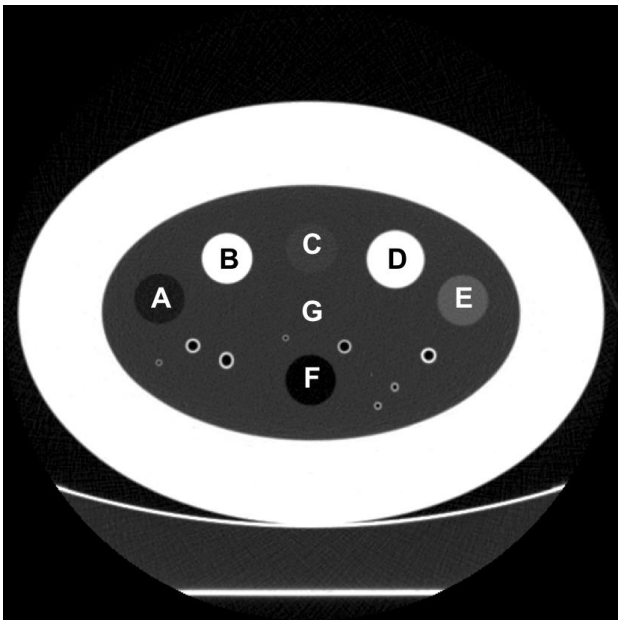

Figure 3. CT slice image of the "COPD Gene" lung phantom used in this study. The phantom consists of (A) 4 lb/ft^3, (B) acrylic, (C) 12 lb/ft^3, (D) water, (E) 20 lb/ft^3, (F) air, and (G) a larger oval shaped lung density equivalent foam surrounded by an outer chest wall equivalent uniformity ring.

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 product 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/m^3) that have been calibrated to their true densities. Such a phantom is introduced and described in [6] and has been shown to establish a HU-electron density relationship for a given scanner and protocol. Related methods[18] may also be considered for harmonization, although this is not strictly necessary to meet the longitudinal claims.

The diagnostic radiology QMP for a site 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 E 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.

4.1.1 Assessment Procedure: HU Bias and Repeatability

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

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. 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 lung equivalent foam 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 acquisitions of the phantom in order to measure the noise and standard deviation of mean HU values in the procedures below.
3. The assessor shall then scan a phantom containing regions of uniform low density such as the COPDGene 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 4.1.2 and 4.1.3.
4. After the scan is performed, the assessor shall place a 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. The assessor shall draw an ROI to avoid the outer 2 mm of the cylindrical material insert but containing at least 3,000 voxels for these measurements. Standard commercial and open source software packages that are sufficient for ROI placement and measurement are published [19], [e.g. http://airwayinspector.acil-bwh.org/].
5. The assessor shall record the values reported for the ROI mean and standard deviation

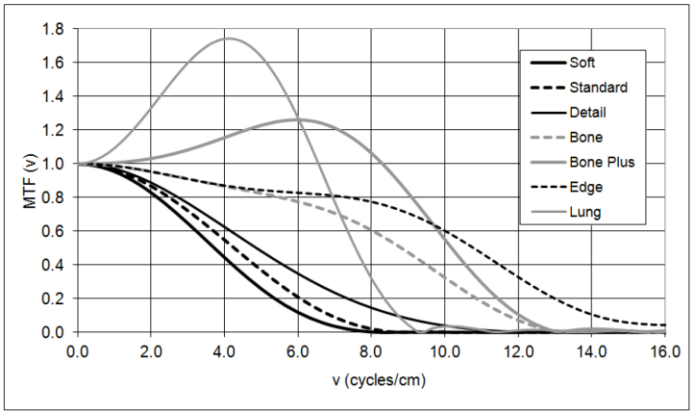

Figure 4: The MTF's for various reconstruction kernels on GE CT scanners. The Standard kernel (bold dashed line) reflects the typically smooth low frequency MTF recommended in qCT densitometry studies.

of the HU. Specifications are restated here for convenience and include:

a. 1000 HU ± 6 HU for inside air (within phantom), and 0 HU ± 6 HU for water (within phantom), and

b. for N = 5 repeated acquisitions, a standard deviation of ≤ 1 HU for inside air (within phantom), lung equivalent foam (within phantom), and water (within phantom).

4.1.2 Assessment Procedure: Voxel Noise

This procedure is recommended for use by the diagnostic radiology QMP for a site or vendors to assess voxel noise. Voxel noise is assessed in terms of noise bias and standard deviation as it applies to lung density analysis, with further references to calculation of the noise power spectrum (NPS) in circumstances where the reconstruction kernel is unknown or needs to be matched to a quantitative reference. 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:

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 (see Table 3.1.2). Specifications are restated here for convenience and include:
    a. noise bias for the mean within the ROI of the subtracted image is ≤ ± 1 HU, and
    b. the standard deviation of the noise within the ROI of the subtracted image is ≤ 20 HU.

Given the proprietary nature of CT reconstruction kernels, some care in comparing reconstruction kernels, or "algorithms", used for reconstruction across vendors is warranted. The assessor shall confirm that the reconstruction kernel for the proposed protocol design (Section 3.4 and Table 3.4.2) is matched to an appropriately smooth modulation transfer function (MTF) such as the GE Standard (bold dashed line in **Figure 4**), Siemens B31f or B34f,

Philips B, and Cannon FC17 kernels.

For multi-center protocols that use different CT vendors in particular, calculation of the MTF and certification by the vendor that all reconstruction kernels are well matched with respect to their MTF is important. Methods for calculating the MTF are published [10,12]. 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.

4.1.3 Assessment Procedure: In-Plane Spatial Resolution and Edge Enhancement

This assessment procedure is used by the diagnostic radiology QMP for a site or vendors to confirm that in-plane spatial resolution specifications are met. The in-plane are assessed in terms of FWHM and require that a FWHM value be computed and recorded. Additionally, edge enhancement is evaluated to ensure that the reconstruction kernel is sufficiently smooth to prevent ringing that can amplify noise and confound quantitative evaluation.

Procedures for assessing PSF and and ERF have been published using standard CT performance phantoms such as the ACR phantom or CATphan. Alternatively, 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] can be used to evaluate spatial resolution using the COPD Gene phantom itself. At present methodology for estimating in-plane and through-plane resolution using the COPD Gene phantom is in development and the committee recommends the ACR and CATphan as the reference standards. .

Using either the ACR or CATphan, the procedure for measuring Edge Enhancement (EE) shall be performed as follows:

1. The assessor shall derive the EE from the edge response function according to the equation:

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

   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%.

The assessor shall use the CATphan or similar; the assessor shall perform the procedure for measuring through-plane resolution (slice profile) as follows:

4.1.4 Assessment Procedure: Through-Plane (Z-axis) Spatial Resolution

This assessment procedure is used by the diagnostic radiology QMP for a site or vendors to

confirm that through plane (axial) spatial resolution specifications are met. The through-plane resolution is assessed in terms of FWHM to evaluate nominal slice thickness, and full width at tenth max (FWTM) to further confirm shape of the slice profile and minimal side-lobe. The procedures require that a FWHM and FWTM value be computed and recorded.

1. The assessor shall assess the through-plane resolution using the slice sensitivity profile with methods described in Fuchs et al. [8] and standard procedures with CT performance phantoms such as the ACR or CATphan phantoms. 4.2. Performance requirements for analysis software.

Given that the analysis software is a source of variability that can be easily controlled by the investigator, the assessor shall use the same analysis software for longitudinal evaluation at multiple time points.

Specifically, the analysis software:

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.

Given that analysis tools are rapidly evolving for quantitative CT of the lungs, the committee recommends a qualifying procedure for new software analysis tools as detailed in Appendices F and G. The CT 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 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 compared in Appendix FF 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 reproducability 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 EF: 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.

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.

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 B.

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

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.

Vendors publishing a QIBA Conformance Statement shall provide a set of "Model-specific Parameters" (as shown in Appendix E) 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

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.
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.
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.
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.
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.
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.
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.
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.

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.
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.

Appendices

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 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.

Appendix B: Conformance Checklists

This Checklist is organized by "Actor" for convenience. If a QIBA Conformance Statement is already available for an actor (e.g. your analysis software), you may choose to provide a copy of that statement rather than confirming each of the requirements in that Actors checklist yourself. Within an Actor Checklist the requirements are grouped by the corresponding Activity in the QIBA Profile document. If you are unsure about the meaning or intent of a requirement, additional details may be available in the Discussion section of the corresponding Activity in the Profile. Conforms (Y/N) indicates whether you have performed the requirement and confirmed conformance. When responding **N**, please explain why. Several of the requirements mandate the use of specific assessment procedures described in Section 4 in the main body of this Profile. Feedback on all aspects of the Profile and associated processes is welcomed.

ACQUISITION DEVICE CHECKLIST

Parameter	Conforms (Y/N)	Requirement
Sample Protocol		Shall prepare a sample protocol conformant with Section 3.4.2, "Protocol Design Specification"
Noise Performance		Shall demonstrate noise bias is ≤ ± 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		Shall demonstrate a Full-width at half-maximum (FWHM) ≤ 1.0 mm as described in Section 4.1.3, Assessment Procedure: In-

		Plane, Through-Plane (Z-axis) Spatial Resolution and Edge Enhancement.
Through-plane spatial resolution		Shall demonstrate a slice sensitivity profile with FWHM ≤ 1.0 mm as described in Section 4.1.3 , Assessment Procedure: In-Plane, Through-Plane (Z-axis) Spatial Resolution and Edge Enhancement.
Edge Enhancement		Shall demonstrate an edge enhancement ≤ 3% for the edge response function as described in Section 4.1.3, Assessment Procedure: In-Plane, Through-Plane (Z-axis) Spatial Resolution and Edge Enhancement.
Acquisition speed		Shall set parameter values that will cover an axial field of view of 35 cm in 10 seconds or less.
Measured HU (Bias)		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) as described in Section 4.1.1, Assessment Procedure: HU Bias and Repeatability.
HU Stability (Repeatability)		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 as described in Section 4.1.1, Assessment Procedure: HU Bias and Repeatability.

IMAGE ANALYSIS SOFTWARE CHECKLIST

Parameter	Conforms (Y/N)	Requirement
Lung Density Analysis		Shall calculate and output for the whole lung: ● RA-950 HU ● Perc15 ● Lung Density Histogram ● Total Lung Volume As described in Section 4.2, Assessment Procedure: Reproducibility of Image Analysis Software across Various Vendors.

Lung Density Analysis		Shall use a consistent lung segmentation procedure including the following steps. 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 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.
Reproducibility of Analysis Software		Shall use identical measurement algorithm for each longitudinal time point measured. See Section 4.2, Assessment Procedure: Reproducibility of Image Analysis Software across Various Vendors for more information.

RADIOLOGIST CHECKLIST

Parameter	**Conforms (Y/N)**	**Specification**
Acquisition Protocol		Shall prepare a protocol to meet the specifications in this table.
Acquisition Protocol		Shall ensure technologists have been trained on the requirements of this profile.

Total Collimation Width		Shall set to Greater than or equal to 16 mm.
Nominal Tomographic Section Thickness		Shall set to Less than or equal to 1.0 mm using procedures in 4.1.3
Scan Duration		Shall set parameter values that will cover an axial field of view of 35 cm in 10 seconds or less.
Reconstruction Protocol		Shall prepare a protocol to meet the specifications in this table.
Reconstruction Protocol		Shall ensure technologists have been trained on the requirements of this profile.
Use of intravenous contrast		Shall not use intravenous contrast.
Use of oral contrast		Shall not use oral contrast.
CT Dose		Shall target less than or equal to 3mGy CTDIvol for a 75kg subject allowing for increased/decreased CT dose adjusted based on patient size and shape according to manufacturer.

PHYSICIST CHECKLIST

Parameter	**Conforms (Y/N)**	**Specification**
Monthly QA		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		Shall, if the acquisition device fails Monthly QA, repeat Product Qualification (See 3.1.2) to re-establish standardization.

Scanner Calibration		Shall assess the CT conformance for the measured HU value and standard deviation in 3.1.2 are met using procedures in 4.1.1 on a monthly basis.
HU Stability		Shall meet the specifications in Table 3.1.2.
Reconstructed Image Thickness		Shall set to 1.0 mm or less.
In-plane Resolution		Shall validate that the protocol achieves a full width at half maximum (FWHM) of line spread function ≤ 1.0 mm using procedures in 4.1.3.
Through-plane spatial resolution		Shall validate that the protocol achieves a slice sensitivity profile with FWHM ≤ 1.0 mm using procedures in 4.1.3.
Edge Enhancement		Shall validate that the protocol achieves a minimum edge enhancement of 3% for the edge response function as described in Section 4.1.3.
Voxel Noise		Shall validate that the protocol achieves a standard deviation of voxel noise that is ≤ 20HU for lung equivalent foam, air and water materials inside a phantom as described in Section 4.1.2.

TECHNOLOGIST CHECKLIST

Parameter	**Conforms (Y/N)**	**Requirement**
Subject Positioning		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		Shall adjust the table height for the mid-axillary plane of the chest to pass through the isocenter.
Subject Alignment		Shall position the subject such that the “sagittal laser line” lies along the sternum (e.g. from the suprasternal notch to the xiphoid process).
Scan Projection Radiograph		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.

Breath-hold Coaching		Shall coach the subject on Breath-holding as described above.
Artifact Sources		Shall remove or position potential sources of artifacts (specifically including breast shields, metal-containing clothing, EKG leads and other metal equipment) such that they will not degrade the reconstructed CT volumes.
Acquisition Protocol Selection		Shall select a protocol that has been previously prepared and validated for this purpose (See section 3.4.2 "Protocol Design Specification").
Acquisition Protocol Selection		Shall report if any parameters are modified beyond the specifications in section 3.4.2 "Protocol Design Specification."
Acquisition Protocol Selection		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.
Scan Plane		Axial / Transverse
Scan Projection Radiograph		Shall confirm the absence of metal or other artifacts
Anatomic Coverage		Shall ensure the Full Lung, from 2cm above the apex to 5cm below the base, is covered by the scan
Axial field of view		Shall confirm the display field of view is no more than 2 cm outside maximal lung extent.
Axial field of view		Shall match the display field of view to that of the Baseline scan, if available.
Adequate coverage		Shall confirm the lung volume is fully represented in the field of view
Motion		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.		Shall confirm the image headers (Single Slice Collimation (0018,9306) and Slice Thickness (0018,0050)) indicate the acquired and reconstructed resolutions ≤ 1 mm
Conformance to baseline.		Shall confirm the Protocol is consistent with the baseline. .
Absence of IV contrast		Shall confirm the absence of IV contrast
Lung Volume		Shall confirm that the proper breath-hold coaching procedure outlined in section 3.5.2 is followed.
Image Distribution		Images shall be transmitted in uncompressed DICOM format and according to the anonymization standards approved for the study.

CLINICIAN OR STATASTICIAN CHECKLIST

Parameter	Conforms (Y/N)	Requirement
Lung Density Analysis		Shall calculate and output the whole lung RA-950, Perc15 and Lung Density Histogram: The analysis software used shall be identical for each longitudinal time point measured (reanalyze images if necessary), 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, Assessment Procedure: Reproducibility of Image Analysis Software across Various Vendors for more detail on procedures

Appendix C: 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 lungs. Only whole lung measurements are considered. Regional and lobar measures of emphysema are increasingly being investigated and reported in the literature [1-6]. 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 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 relevant [11]. This will 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 1 [12] as the measures best correlated to microscopic histology, and Perc15 [4] 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.
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 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.

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 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 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 [22]that was included as a reference example in an "asymptomatic 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. **<u>Effect of VA on Bias and Precision Claims for RA-950 and Perc15</u>**

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 guidelines [38]. QIBA statistical methods guidelines [39] and terminology guidelines [40] 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 model [38]. 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 assessments [41-43]. 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 study [22], 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 definitve claim when using the RA-950 metric in combination with VA when more data becomes available.

5. <u>Claims:</u>

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 [Δx -18 HU, Δx +18 HU] with 95% confidence; for a measured change of Δy% in RA-950, one can expect the true change to lie in the interval [Δy -3.7%, Δy +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 [Δx -11 HU, Δx +11 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.) [♭]	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
*[o]MILD trial only (16 slice)
** 140 kVp (vs. 120 kVp) for > 80kg weight
[♮]Smoothing operation in image domain
[♭] 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	ΔLOA	% Δ	Bias (SD)	LOA	Bias (SD)	LOA	ΔLOA	% 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
Hochhegger	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).

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

[2]linear 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

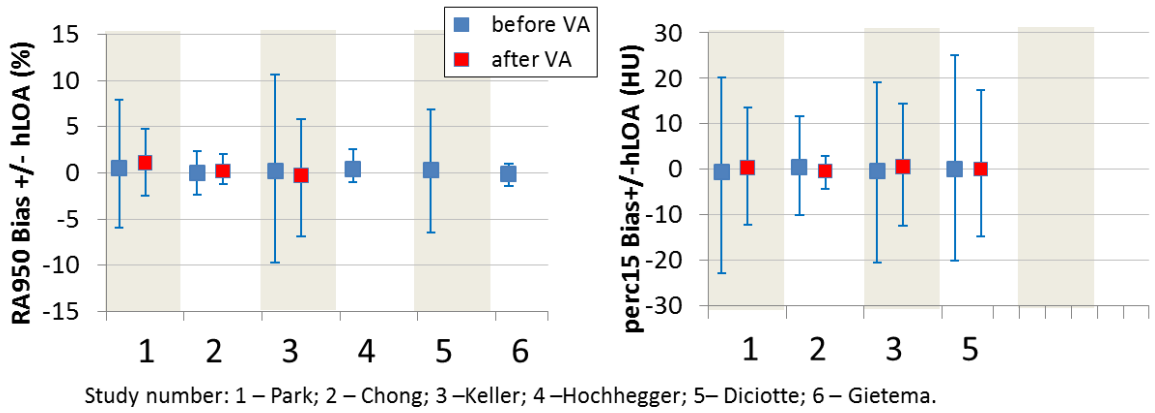

Fig. 1. Bias and limits of agreement compiled from the selected 6 studies: Bias and 95% limits of agreement (hLOA, or half-width of the 95% LOA, defined as 1.96 SD_{bias}) for RA-950 HU (left panel) and Perc15 (right panel) both before (blue) and after (red) VA for selected studies (Table 1). The respective 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.

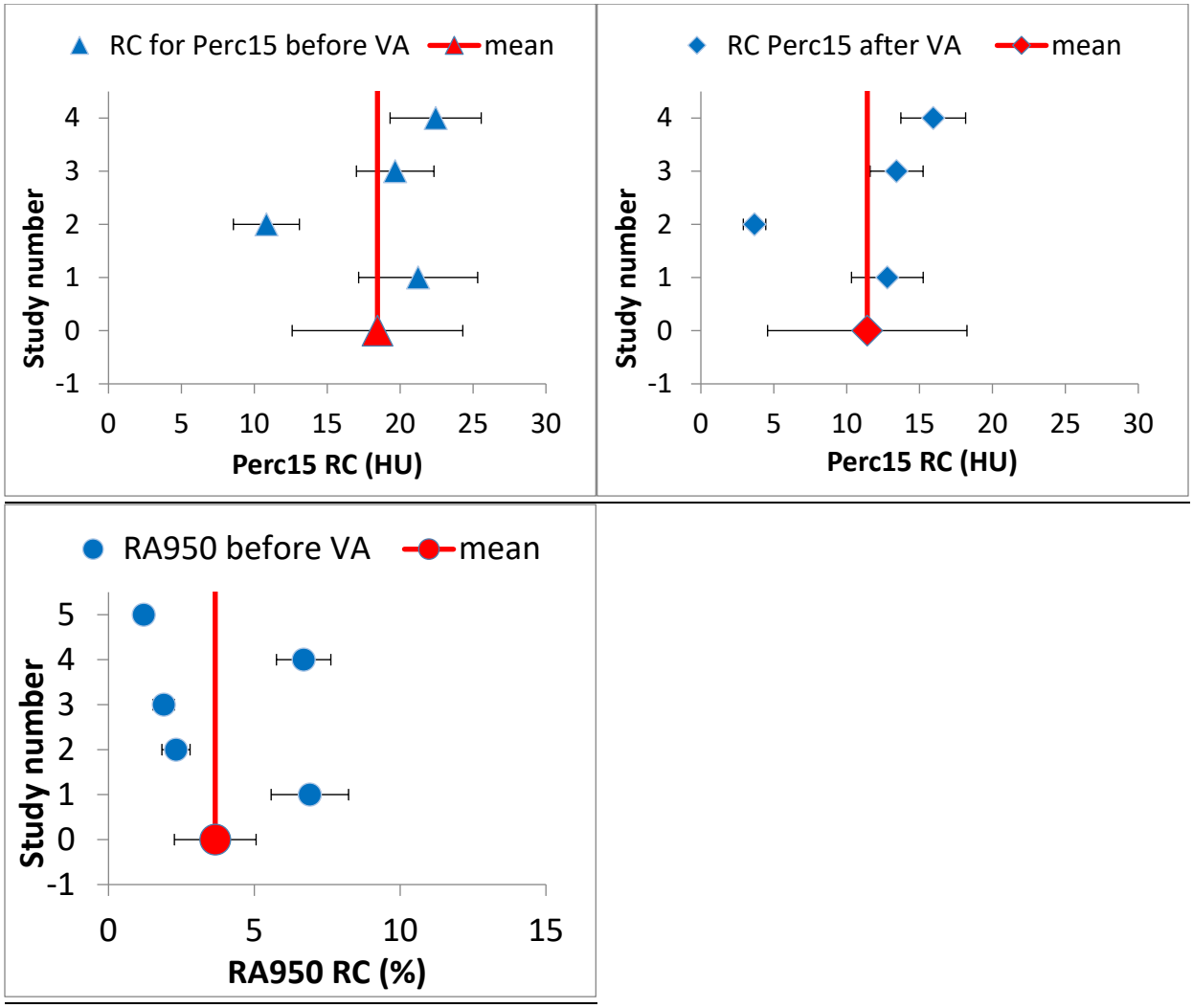

Fig 2. Forest plots of the repeatability coefficient (RC) of the studies (blue) and their pooled RC (red), with the 95% CI (defined as 1.96 √SE_RC) shown as error bars: (a) Perc15 before VA, (b) Perc15 after VA, (c) RA-950 before VA. For RA-950, only two studies performed VA, and the effect of VA on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

Appendix D: 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 for [40]. 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 as [38] 1.96 √2 wSD. Based on the defintions, 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

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 itself [37]. The variance of RC^2 is $Var(RC^2)=kb^2$, where $k=N(P-1)/2$ is the shape parameter (P 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

The goal of the meta-analysis is to derive a mean RC based on the four studies selected. Following [37], 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 that the random effects model be used. A variance τ^2 representing the underlying distribution of the RC's is constructed based on the DerSimonian and Laird estimator for this group of studies [37], 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 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 E: Model-specific Instructions and Parameters

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.

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.

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.

Appendix F: Image Analysis Software Comparison

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

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 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:

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

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.

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.

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

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

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

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

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

vendor l.

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

The average RDC for vendor l is then given by:

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

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.

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.

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

	Inter-software RDC Without QA		Inter-software RDC With QA	
Parameter	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 G. Assessment procedure for computing reproducibility of Lung Density Software across various vendors

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.

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.

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). 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 F.

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 ith image set from the 50 subjects.

The RDC is defined according to section 7.2 of Obuchowski et al [45] as:

$$RDC = 1.96 * \sqrt{2\sigma^2_{inter-software}}$$

$$\sigma^2_{inter-software} = \frac{1}{N-1}\sum_{i=1}^{N}(M_i - \bar{M}_i)^2$$

$$\bar{M}_i = \frac{1}{K}\sum_{k=1}^{K} M_k$$

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 ith image set and kth algorithm.

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.

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 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:

$$RDC_{overall} = 1.96 * \sqrt{2\left(wSD^2 + \sigma^2_{inter-software}\right)}$$

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

References for Appendices A-G:

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.
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.
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.
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.
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.
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.
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.
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.
16. Hochhegger B, Irion KL, Marchiori E, Moreira JS. Reconstruction algorithms and their influence in emphysema CT measurements. *Acad Radiol.* 2010;17(5):674.

17. Hochhegger 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.
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.
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.
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.
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.
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.
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.
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.
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.

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.
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.
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.
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.
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.
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.
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.