Reference standard and statistical model for intersite and temporal comparisons of CT attenuation in a multicenter quantitative lung study

J. P. Sieren and J. D. Newell, Jr.

Department of Radiology, University of Iowa Carver College of Medicine, Iowa City, Iowa 52242

P. F. Judy

Department of Radiology, Brigham and Women's Hospital, Boston, Massachusetts 02115

D. A. Lynch

Department of Radiology, National Jewish Health Center, Denver, Colorado 80206

K. S. Chan

Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa 52242

J. Guo and E. A. Hoffman^{a)}

Department of Radiology, University of Iowa Carver College of Medicine, Iowa City, Iowa 52242

(Received 18 February 2012; revised 10 July 2012; accepted for publication 7 August 2012; published 4 September 2012)

Purpose: The purpose of this study was to detect and analyze anomalies between a large number of computed tomography (CT) scanners, tracked over time, utilized to collect human pulmonary CT data for a national multicenter study: chronic obstructive pulmonary disease genetic epidemiology study (COPDGene).

Methods: A custom designed CT reference standard “*Test Object*” has been developed to evaluate the relevant differences in CT attenuation between CT scanners in COPDGene. The materials used in the *Test Object* to assess CT scanner accuracy and precision included lung equivalent foam (−856 HU), internal air (−1000 HU), water (0 HU), and acrylic (120 HU). Nineteen examples of the *Test Object* were manufactured. Initially, all *Test Objects* were scanned on the same CT scanner before the *Test Objects* were sent to the 20 specific sites and 42 individual CT scanners that were used in the study. The *Test Objects* were scanned over 17 months while the COPDGene study continued to recruit subjects. A mixed linear effect statistical analysis of the CT scans on the 19 *Test Objects* was performed. The statistical model reflected influence of reconstruction kernels, tube current, individual *Test Objects*, CT scanner models, and temporal consistency on CT attenuation.

Results: Depending on the *Test Object* material, there were significant differences between reconstruction kernels, tube current, individual *Test Objects*, CT scanner models, and temporal consistency. The two *Test Object* materials of most interest were lung equivalent foam and internal air. With lung equivalent foam, there were significant ($p < 0.05$) differences between the Siemens B31 (−856.6, ± 0.82 ; mean $\pm$ SE) and the GE Standard (−856.6 $\pm$ 0.83) reconstruction kernel relative to the Siemens B35 reference standard (−852.5 $\pm$ 1.4). Comparing lung equivalent foam attenuation there were also significant differences between CT scanner models ($p < 0.01$), tube current ($p < 0.005$), and in temporal consistency ($p < 0.005$) at individual sites. However, there were no significant effects measurable using different examples of the *Test Objects* at the various sites compared to the reference scans of the 19 *Test Objects*. For internal air, significant ($p < 0.005$) differences were found between all reconstruction kernels (Siemens B31, GE Standard, and Phillips B) compared to the reference standard. There were significant differences between CT models ($p < 0.005$), and tube current ($p < 0.005$). There were no significant effects measurable using different examples of the *Test Objects* at the various sites compared to the reference scans of the 19 *Test Objects*. Differences, across scanners, between external air and internal air measures in this simple (relative to the *in vivo* lung) test object varied by as much as 15 HU.

Conclusions: The authors conclude that the *Test Object* designed for this study was able to detect significant effects regarding individual CT scanners that altered the CT attenuation measurements relevant to the study that are used to determine lung density. Through an understanding of individual scanners, the *Test Object* analysis can be used to detect anomalies in an individual CT scanner and to statistically model out scanner differences and individual scanner changes over time in a large multicenter trial. © 2012 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4747342>]

Key words: chronic obstructive pulmonary disease (COPD), emphysema, quantitative CT, lung density, multicenter trial

I. INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a progressive condition affecting an estimated 24 million individuals in the United States¹ and is the fourth leading cause of death in the United States.² Most patients with COPD acquire the disease through cigarette smoking. However, it is estimated that 25%–45% of patients with COPD are never-smokers especially in developing countries and have acquired the disease by exposure to the burning of biomass fuel, occupational exposure to dusts and gases, chronic asthma, pulmonary tuberculosis, respiratory-tract infections during childhood, and outdoor air pollution.^{3,4} This means that worldwide COPD is a serious medical condition for both the nonsmoker and current or former smokers. Also, not all smokers develop COPD.⁵ This suggests that COPD has a substantial genetic component and that some individuals are at increased risk of developing COPD.⁶

A commonly used method for classifying the stage of COPD is the global initiative for chronic obstructive lung disease (GOLD) method. Unfortunately, the GOLD staging system fails to classify patients accurately according to disease subtype and disease severity.⁷ Objective, quantitative methods for assessing CT images have emerged to track presence, distribution, and progression of pathologic processes in subjects at risk for COPD. These measures include assessment of emphysema-like lung regions, regional air trapping, and measures of changes in airway geometry from disease induced remodeling.^{8–10} The COPD genetic epidemiology study (COPDGene) was designed to identify those patients with increased genetic risk of developing COPD.¹¹ CT evaluation is a critical component of this study, to quantify the severity and extent of emphysema, air trapping, and airway wall thickening. The study has enrolled 10 000 subjects and obtained an inspiratory and an expiratory chest CT on each of these participants. The large number of subjects is required because of the genome analysis techniques being used in the study.¹¹ In order to recruit this large number of subjects it was necessary to have 20 clinical sites enrolling patients. The CT scanners at these sites varied by manufacturer and model so for this reason, it was necessary to devise a method to determine the level of variation in CT lung attenuation measurements between the CT scanners in the study, and the longitudinal variation in measurement for individual scanners. Comparable CT scanning protocols for the models of scanners used during the study were specified (Table I). The purpose of this paper is to describe the design and results using the purpose built *Test Object* to assess anomalies between the many CT scanners involved in the COPDGene and the temporal stability of each scanner. While the *Test Object* allowed for the identification of large step-wise shifts in scanner calibration between two discrete time points, these step-wise shifts are not the focus of the presented statistical model but rather the characterization of inherent baseline differences between scanners and their drift through the course of the study. Furthermore, it is recognized that *in vivo* scanning brings out added effects, not fully accounted for by the developed *Test Object* including differences in the treatment

TABLE I. CT scanner models used to acquire data for the COPDGene study.

Scanner models used in the COPDGene study		
Vendor	Model	Number of scanners
Siemens	Biograph 40	1
Siemens	Sensation 16	4
Siemens	Sensation 64	4
Siemens	Definition	5
Siemens	Definition AS+	1
Siemens	Definition Flash	1
GE	Light Speed 16	6
GE	Light Speed 16 Pro	1
GE	Light Speed VCT	10
GE	Discovery CT750HD	2
Philips	Brilliance 40	1
Philips	Brilliance 64	6

of scatter and beam hardening, known to affect reconstructed CT Hounsfield units (HU).

II. MATERIALS AND METHODS

II.A. COPDGene test object

The custom designed COPDGene *Test Object* (manufactured to specifications by the Phantom Laboratories, Salem, NY) (Fig. 1) consists of an outer ring that has the x-ray attenuation characteristic similar to water (7–20 HU). The inside of the ring contains a foam material with an average CT attenuation close to the average attenuation of normal lung parenchyma (–856 HU). The schematics of the *Test Object* [Fig. 2(a)], demonstrates that the lung-equivalent foam contains 12 internal holes and six embedded polycarbonate tubes [Fig. 2(b)]. Two of the tubes were embedded at a 30° angle relative to nominal axis of the *Test Object*. The tubes are used to assess CT measurements of airway geometry. However, the airway geometry results are not the subject of this paper. Three other materials are embedded within the foam material. A 3.2 cm polypropylene bottle filled with distilled water, a 3 cm air hole, and a 3 cm acrylic rod. All 19 *Test Objects* were first scanned on a single CT scanner

FIG. 1. Image of the *Test Object*.

FIG. 2. (a) Schematic of the COPDGene *Test Object* including the content and dimensions of the internal water, acrylic, and air insert structures. Adopted from final Phantom Lab's schematics for "CTP657." (b) Schematics of the airway tubes, internal, and external dimensions. The specific geometry of the *Test Object* was used in design of the automated analysis software. Adopted from final Phantom Lab's schematics for "CTP657."

(Siemens Sensation 64, Iowa Comprehensive Lung Imaging Center (I-CLIC), University of Iowa) to provide the ability to examine the consistency of the CT attenuation measurements in each of the 19 examples of the *Test Object* before these *Test Objects* were disseminated to the COPDGene study sites.

II.B. COPDGene study sites and scanner models

The COPDGene sites were supplied with a specific, identifiable *Test Object*. Two sites within close geographic location shared a single *Test Object*, resulting in 19 *Test Objects* for 20 sites. Sites were instructed to scan the *Test Object* monthly using the COPDGene human subject protocol, which included two different tube currents, 200 mAs (inspiratory scan) and 50 mAs (expiratory scan). The COPDGene human protocol standardized on a kv of 120 while slice thickness (0.6–0.9 mm) and pitch (0.984–1.375) were chosen to be similar across scanners within the limits of what was possible for a given scanner. Scanned field of view was standardized across models to use the maximum beam width for the largest subjects. The phantoms were thus scanned at each site

FIG. 3. (a) Schematic illustrating the automated segmentation of the tube inserts, demonstrating avoidance of partial volume influence by excluding the outer boundaries of the insert. (b) Schematic of the simulated parenchyma segmentation, excluding the influence of the airway tubes.

to match the human protocol at that site. Because these variables were fixed for the COPDGene study, we do not evaluate within scanner variability of these parameters as sources of quantitative data anomalies in this paper. While the diameter field of view (DFOV) varied amongst human subjects so as to reconstruct tightly to the rib cage, DFOV for the phantom was standardized at 36.5 cm. The University of Iowa developed and distributed a written instruction manual to demonstrate proper positioning of the *Test Object* within the scanner to ensure that data were collected in a standardized way. During a 17-month period, scans were collected from 42 CT scanners used in the COPDGene study, including 12 different models from three manufacturers (Table I).

II.C. Test object analysis and segmentation

Test Object CT data were sent to the Iowa Comprehensive Lung Imaging Center (I-CLIC), for detailed analysis. The quality control system consisted of a protocol adherence check prior to accepting the CT scans. This process removed unwanted scans from the dataset and gave notice to a given site that there was a mistake during the CT acquisition. The accepted scans were assessed by an automated software module of pulmonary analysis software suite (PASS) (Ref. 12) which extracted consistent measurements of uniform regions of interest (ROIs) in HU for lung-equivalent foam, internal air, external air, water, and acrylic material.

The 30 mm in diameter cylinders of air, water, and acrylic regions [Fig. 3(a)] were segmented by a thresholding method

followed by connected component analysis. The segmented regions were eroded by four pixels around the outer edge to eliminate the partial volume effect at the boundary. The segmented depth (z axis) was 20 mm. The same approach was applied to segment the elliptically shaped lung equivalent foam material, [Fig. 3(b)]. External air was sampled by using a 30 mm $\times$ 30 mm square ROI with a z-axis depth of 20 mm in the top center of the air-containing region outside of the *Test Object*.

II.D. Statistical analysis methods for the COPDGene test object

The image data from the 20 centers were separated into the 42 individual CT scanners used in the study with each CT scanner given a CT ID number. One of the 19 *Test Objects* was to be scanned each month by each CT scanner in the study over a period of 17 months. Unfortunately, the *Test Object* was not scanned each month at all sites. In some instances it was only scanned once for the initial acceptance scan by a particular CT scanner. This resulted in a sparse table of data over time, in part influencing the design of a more complex statistical modeling approach known as the linear mixed effect model.¹³ The statistical analysis was based on tracking each of these individual CT *Test Objects* to the particular date and CT scanner where each of the CT *Test Objects* was subsequently scanned.

We employ the linear mixed model framework for data analysis, mainly because variables of interest (reconstruction kernels, tube current, CT scanner models, and temporal consistency) contain numerous distinct values whose particular effects on the response are not of primary importance. For instance, there are 19 distinct *Test Objects* such that a standard regression analysis requires suppressing the intercept and introducing dummy variables for each of the 19 *Test Objects*. However, the specific, *fixed effects* of each particular *Test Object* on the response are infeasible to estimate due to an inadequate sample size. Instead, the *random effects* approach assumes the coefficients of the 19 dummy variables collectively constitute a *random* sample from some normal distribution with unknown mean and unknown standard deviation that quantifies the test-object-to-object variation in their effects on the response. If the standard deviation equals zero, the *Test Objects* are identical and do not contribute to the variations in the response. Hence, the significance of the *Test Objects* can be tested by checking whether or not the standard deviation equals 0, which can be assessed by a likelihood ratio test. Often, the coefficients are decomposed as the sum of an overall mean plus deviations where the deviations are then a random sample from a normal distribution with zero mean and unknown standard deviation; strictly speaking the deviations are the random effects. The random effects approach greatly reduces the number of parameters (from 19 to 2 for modeling *Test Objects*), resulting in a more parsimonious model. Time consistency data from each scanner we simply model as a linear function of time, with the initial scan date being the origin of time, and whose slope and intercept are modeled as a random vector from a bivariate normal distribution. Similarly, CT

scanner model is modeled by the random effects approach. The impact of all other covariates, including reconstruction kernel, tube current, and the reference standard (Siemens Sensation 64), used as a dummy variable, are modeled as fixed effects. If the coefficient estimate of the latter dummy variable is not significantly different from zero, then all scanners are identical to the reference standard (Siemens Sensation 64), on average, after adjusting for difference in other covariates (reconstruction kernel, tube current, *Test Objects*, CT scanner model, and time consistency).

A separate linear mixed effect model was fitted for each of the five materials of interest in the *Test Object* that included lung equivalent foam, internal air, external air, water, and acrylic. The general linear mixed effect model approach looks at all fixed effects and all random effects involved in the study and through an iterative process reduce the number of variables in the model until it converges using the likelihood ratio test to exclude insignificant variables. Variables modeled using the fixed effect approach included the reconstruction kernel, and tube current. The variables modeled via the random effects approach included *Test Objects*, CT scanner model, and time consistency of the intercept, and slope of each CT scanner used. Statistical significance is based on 5% significance value, i.e., p value less than 0.05 for either fixed effect or random effect. A random effect is said to be significant if the standard deviation of the random effect is significantly different from zero, at 5% significance level (p -value < 0.05). For example, if the standard deviation of the *Test Object* effects for lung equivalent foam measurements is significantly different from zero at the 5% level, then HU varies with *Test Objects*. On the other hand, if the standard deviation is zero, then HU does not vary with *Test Objects*. So, the individual *Test Object* has (no) association with HU of lung equivalent foam or the variability of an individual *Test Object* adds (no) explanation of HU in lung equivalent foam when p -value is less (greater) than 0.05. Other random effects can be similarly interpreted.

II.E. The linear mixed model

The data for a particular response, such as lung equivalent foam, are grouped by *Test Object*, then by CT scanner model, and finally by individual scanner. In symbols, let $y_{i,j,k,t}$ denote the measurement of the i th *Test Object* scanned by a scanner of model j , specifically by the k th scanner on the t th day since the date of initial scanning by that scanner. Note the range of j is specific to *Test Object* i , and the range of k is specific to the combination of *Test Object* i ($i = 1, 2, \dots, 19$) and model j , while the range of t is specific to the *Test Object* i scanned by scanner k . For each measurement, there is an associated covariate vector $x_{i,j,k,t}$ whose components are the values of the following seven dummy variables, namely, Siemens B35, Siemens B31, Philips B, and GE Standard reconstruction kernels, the reference standard CT scanner (Siemens Sensation 64), tube currents 50 mAs vs 200 mAs (being equal to 1 for 200 mAs and 0 otherwise). The linear mixed model fitted to each of the five measurements takes the

general form:

$$y_{i,j,k,t} = x_{i,j,k,l}^T \beta + b_i + b_{i,j} + b_{i,j,k,0} + t \times b_{i,j,k,1} + \epsilon_{i,j,k,t}, \quad (1)$$

where the superscript T denotes taking the transpose of the vector, the coefficient vector β describes the fixed effects due to reconstruction kernel, tube current and difference from the reference standard scanner; the latent variables b_i model the *Test Object* effects and are assumed to be independent and identically (iid) normally distributed with zero mean and standard deviation $\sigma_{\text{test object}}$; the random variables $b_{i,j}$ attempt to capture the latent scanner model effect for *Test Object* i and are iid, centered normal variables (i.e., with zero mean) of standard deviation σ_{model} ; $b_{i,j,k,0}$ model the scanner variation in the initial scan and are iid, centered normal with standard deviation σ_0 ; $b_{i,j,k,1}$ model the scanner variation in the slope of the linear time consistency trend and are iid centered normal variables of standard deviation σ_{slope} ; the last two normal variables are allowed to be correlated, but otherwise the b 's are uncorrelated with each other. The b 's are the random effects of the model. Finally, the terms $\epsilon_{i,j,k,t}$ are also normally distributed error variables with zero mean and standard deviation σ_{error} , and the error terms are independent of each other and other random effects, across measurements. See Ref. 13 for a comprehensive survey of the linear mixed model. All linear mixed models are estimated by the method of restricted maximum likelihood (REML). See the Appendix for the implied covariance structure of the model.

II.F. Correlation of test object CT measurements with COPDGene human subject CT data

During the course of the study, the *Test Object* CT scans identified four CT scanners that showed significant shifts in the measured CT attenuation of internal air that was confirmed by using the recommended manufacturer performed CT air calibration. This prompted further investigation to determine if these CT attenuation shifts were also measurable on COPDGene subject CT scans acquired during the same period. Twenty-four human study scans were obtained within 4 weeks before and after scanner calibration on scanner 12. Human tracheal luminal air data were analyzed fully automatically (Apollo 1.0, VIDA Diagnostics, Iowa City, IA). The airway tree lumen were identified using a 3D region grow based algorithm. The centerlines of all airway segments were extracted, and the airway branch points were identified. The segmented regions were eroded to eliminate the partial volume effect at the boundary. Only the lower third (closest to carina) was used for further evaluation to avoid the influence of the increased occurrence of CT noise typically observed in the upper part of the trachea. The median CT density of the lower third trachea voxels was then reported for further evaluation since the median CT attenuation of tracheal air, best human equivalent for *Test Object* internal air, was measured before and after the *Test Object* shifts occurred. Tracheal air was chosen because it is a large sample of air inside the human

body. An unpaired t -test was used to compare values before and after calibration.

III. RESULTS

CT measurements of the lung equivalent foam, internal air, external air, water, and acrylic materials obtained by scanning 19 different *Test Objects* on 42 individual study CT scanners and the corresponding statistical analysis of these measurements are summarized in Tables II–IV. Important aspects of the statistical analysis using the mixed linear effects model on the data are expanded below for the five materials that were studied.

Data from each scanner generally fluctuated over time, owing to three possible reasons: (1) different *Test Objects* were scanned over time, (2) operation variations due to varying environmental conditions, and (3) sporadic technical problems, e.g., machine malfunctions or recalibration problems. These diverse sources of variation, plus the differences in scanner models, make it hard to compare time-series measurements visually. The linear mixed model has properly accounted for the variations in the *Test Objects*, while environmental effects on the operation of the scanners are likely negligible. However, the presence of technical problems may invalidate the measurements, so it is pivotal to detect their occurrences. Technical problems generally cause nonstationarity in the data, by introducing outliers, level shifts, time inconsistencies and/or changing variance over time (volatility). Here, we simply check possible nonstationarity in the time consistency measurements from each scanner by including a linear time consistency for each scanner in the model; scanners with an atypical time consistency are then deemed as outliers. The linear mixed model specifies that the intercept (level at initial date of scan) and slope of each time consistency trend (average daily rate of change) follow a bivariate normal distribution. A scanner is deemed an outlier, with respect to certain measurement, if it has an extreme combination of intercept and slope, based on its Mahalanobis distance from the center of the bivariate normal distribution, with 5% family error rate. Outliers were then omitted from the final models, because the presence of such outlier(s) suggests that the corresponding scanner(s) are likely incorrectly calibrated in which case data from such scanners may need to be revised, with scanning redone; hence, it is of greater scientific interest to estimate the typical variation in data from “working” scanners. For the results reported below, the estimates are generally robust to the outliers except for the random effects of the intercept and/or the slope of the time trend that generally has a larger variance with the inclusion of the outliers.

III.A. Lung equivalent foam

Site 12 is found to be an outlier as it has an outlying time consistency (based on the procedure described above), so it is excluded from the analysis.

TABLE II. Summary of linear mixed effect model for the fixed effects which included the tube current and reference standard vs individual CT scanners of all five materials studied. There were significant differences in the tube current for lung foam, internal air and water. Mean for all scanners are shown in the middle column while the range of individual scanner differences associated with a change in mAs are shown in the far right column.

Material	Fixed effect: Tube current (200 vs 50 mAs)	
	Mean, standard error	Difference range absolute HU (min–max)
Lung foam	0.3089, $\pm 0.08^a$	0–6
Internal air	0.2866, $\pm 0.07^a$	0–5
External air	0.2430, $\pm 0.08^a$	0–7
Water	0.4410, $\pm 0.14^a$	0–10
Acrylic	–0.0408, ± 0.13	0–9

^a $p < 0.005$.

III.A.1. Fixed effects

The B31 kernel ($p = 0.011$) and the standard kernel ($p = 0.013$) were significantly different from the reference B35 kernel. There was no significant difference between the B kernel ($p = 0.18$) and the B35 reference kernel. There were statistically significant differences between the 200 mAs tube current and the 50 mAs tube current ($p < 0.0001$) (Table II). There was no significant difference between the individual CT scanners ($p = 0.93$) and the reference standard scanner. Mean HU values of the B31 kernel (–856.6 HU) and standard kernel (–856.6) were lower than those of the B35 kernel (–852.5) and B kernel (–854.9). Standard errors of the means values were greater for the B35 kernel (± 1.4) than compared to the B31 (± 0.82), standard (± 0.83) and B (± 0.91) kernels (Table III).

III.A.2. Random effects

For lung equivalent foam, individual *Test Objects* have no association with HU ($p = 0.29$), where the p -value is from the likelihood ratio test that the random effects of the *Test Objects* (b_i , $i = 1, 2, \dots, 19$) are identically zero, i.e., of zero standard deviation. (Below, we shall simply characterize a random effect as significant or not significant depending on whether or not its p -value is less than 5%.) There were significant differences in time consistency ($p < 0.0001$) and between

TABLE IV. Summary of linear mixed effect model for the random effects which included the 19 different *Test Objects*, differences across all scanners, differences within a single scanner model and differences across time. Interestingly, the air, water and acrylic measures demonstrated higher variability within a given scanner model, compared to the variability seen in the same measurements across all scanners in the study.

Material	Random effects (standard deviation, SD)			
	Differences Between Test Objects (SD of b_i)	Differences across all scanners (SD of $b_{i,j}$)	Differences within a scanner model (SD of $b_{i,j,k,0}$)	Differences of time consistency (SD of $b_{i,j,k,1}$)
Lung foam	1.46	2.38 ^a	1.36 ^b	0.01 ^a
Internal air	1.52	0.07 ^a	2.37 ^a	0.01 ^a
External air	—	1.13 ^a	1.12 ^b	0.001 ^a
Water	1.36 ^c	0.64	3.23 ^a	0.008 ^a
Acrylic	0.0004 ^a	1.28 ^c	3.53 ^b	0.004 ^a

^a $p < 0.005$;

^b $p < 0.01$;

^c $p < 0.05$.

CT scanner models ($p = 0.046$) of the individual CT scanner lung foam measurements. Standard deviation between the individual *Test Objects* (1.46) was lower than that which was observed across all scanners (2.38). Standard deviation was higher across all scanners (2.38) than within a CT scanner model (1.36). Standard deviation of the time consistency measurement (0.01) demonstrated slight variability (Table IV).

In order to validate the somewhat complex random-effects specification of the preceding linear mixed model (referred to as the final model in this paragraph), we have also fitted two simpler models with identical fixed-effects specification but with simpler random effects specification: (1) just the random effects of *Test Objects* and (2) just the random effects of *Centers*. The AIC of the simpler models are respectively 3570.2 and 3380.3, much higher than 2956.1, the AIC of the final model. (Note that AIC is defined as minus twice maximum likelihood plus twice the number of parameters,¹³ so smaller AIC is better.) Similarly, the BIC of the simpler models are respectively 3608.3 and 3418.4, as compared to 3008.5, that of the final model. Moreover, the standard deviation of the *Test Objects* in the first simplified model equals 4.4, much higher than 1.46, that of the final model. (The formula of BIC is similar to that of AIC except that the penalty term is the

TABLE III. Summary of linear mixed effect model for the fixed effects (kernels) of all five materials studied. The B35 kernel generally demonstrated larger bias in mean value and more variability in terms of standard error as compared to the other kernels.

Material	Fixed effect: Reconstruction kernel (mean HU, standard error)			
	Siemens B35	Siemens B31	GE Standard	Philips B
Lung foam	–852.5, $\pm 1.40^a$	–856.6, $\pm 0.82^b$	–856.6, $\pm 0.83^b$	–854.9, ± 0.91
Internal air	–997.4 $\pm 1.28^a$	–1004.6, $\pm 0.88^a$	–1004.6, $\pm 0.89^a$	–1002.7, $\pm 0.89^a$
External air	–1001.3 $\pm 0.98^a$	–1000.3, ± 0.54	–1000.3, ± 0.56	–997.5, $\pm 0.73^c$
Water	–5.6, $\pm 1.60^a$	–1.8, $\pm 1.04^b$	–2.6, ± 1.07	–1.1, $\pm 1.05^a$
Acrylic	118.4, $\pm 1.61^a$	122.2, $\pm 0.94^b$	121.2, ± 0.97	118.8, ± 1.01

^a $p < 0.005$;

^b $p < 0.05$;

^c $p < 0.01$.

product of the number of parameters times the logarithmic sample size, again smaller BIC is preferred.) These results lend support to the adopted random-effects specification.

III.B. Internal air

Sites 12 and 37 are found to be outliers and hence removed from the analysis.

III.B.1. Fixed effects

The B31 kernel ($p < 0.0001$), standard kernel ($p < 0.0001$) and B kernel ($p = 0.0043$) were significantly different from the reference B35 kernel. There was a statistically significant difference between the 200 mAs and the 50 mAs tube current ($p < 0.0001$) (Table II). There was no significant difference between the individual CT scanners ($p < 0.51$) and the baseline scanner. Mean HU values of the B31 kernel (-1004.6 HU) and standard kernel (-1004.6) were lower than those of the B35 kernel (-997.4) and B kernel (-1002.7). Standard errors of the means values were greater for the B35 kernel (± 1.28) than compared to the B31 (± 0.88), Standard (± 0.89), and B (± 0.89) kernels (Table III).

III.B.2. Random effects

There were significant differences between individual *Test Objects* ($p = 0.02$), CT models ($p < 0.0001$), and time consistency ($p < 0.0001$) of the individual CT scanner measurements. Standard deviation between the individual *Test Objects* (1.52) was lower than that which was observed across all scanners (2.38). Interestingly, the standard deviation was lower across all scanners (0.07) than within a single scanner model (2.37). Standard deviation of the time consistency measurement (0.01) demonstrated slight variability (Table IV).

III.C. External air

Sites 12, 29 and 37, 38 were eliminated from analysis as stated above.

III.C.1. Fixed effects

There was a significant difference between the B kernel ($p < 0.01$) and the B35 reference kernel. The B31 kernel ($p < 0.40$) and the standard kernel ($p < 0.39$) were not significantly different from the reference B35 kernel. There were statistically significant differences between the 200 mAs and the 50 mAs tube current ($p < 0.036$) (Table II). There was no significant difference between the individual CT scanners ($p < 0.64$) and the baseline scanner. Mean HU values of the B31 kernel (-1000.3 HU) and standard kernel (-1000.3 HU) were higher than the B35 kernel (-1001.3 HU) and slightly lower than the B kernel (-997.5 HU). Standard errors of the means values were greater for the B35 kernel (± 0.98) than compared to the B31 (± 0.54), Standard (± 0.56), and B (± 0.73) kernels (Table III).

FIG. 4. Relationship between inside vs outside air HU values assessed from all *Test Object* scans gathered across sites.

III.C.2. Random effects

As demonstrated in Table III, a model including the variable *Test Object* could not be fitted (algorithm did not converge), which may reflect the fact that the correlation between the attenuation values of internal and external air is 0.48, but the Kendall correlation is 0.066 (only marginally significant at 5% level) and the Spearman correlation is 0.061 (not significant at 5% level). Indeed, the relationship is non-normal, as shown by Fig. 4. There were (marginal) significant differences between CT models ($p = 0.059$) and time consistency ($p < 0.0001$) of the individual CT scanner measurements. Standard deviation was almost the same across all scanners (1.13) than within a scanner model (1.12). Standard deviation of the time consistency measurement (0.001) demonstrated very little variability (Table IV). One should keep in mind that three Light Speed 16 scanners (included in Fig. 4) were removed from other analyses because of large, temporal step changes.

III.D. Water

Site 15 is an outlier, so removed from the analysis.

III.D.1. Fixed effects

There were no significant differences between the standard kernel ($p = 0.13$) compared to the B35 reference kernel. There were significant differences between the B31 kernel ($p = 0.048$) and the B kernel ($p = 0.037$) compared to the B35 reference kernel. There were statistically significant differences between the 200 mAs and the 50 mAs tube current ($p < 0.0012$) (Table II). There was no significant difference between the individual CT scanners ($p < 0.59$) and the baseline scanner. Mean HU values of the B31 kernel (-1.8 HU) and standard kernel (-2.6 HU) were higher than the B35 kernel (-5.6 HU) and slightly lower than the B kernel (-1.1 HU). Standard errors of the means values were greater for the B35 kernel (± 1.6) than compared to the B31 (± 1.04), Standard (± 1.07) and B (± 1.05) kernels (Table III).

III.D.2. Random effects

There were significant differences between individual *Test Objects* ($p < 0.0087$), CT models ($p < 0.0044$), and time consistencies ($p < 0.0001$) of the individual CT scanner measurements. Standard deviation between the individual *Test Objects* (1.36) was higher than that which was observed across all scanners (0.64). Differences within a given model were higher ($SD = 3.23$) than across all the scanners ($SD = 0.64$) for the water measurements. Standard deviation of the time consistency measurement (0.008) demonstrated slight variability (Table IV).

III.E. Acrylic

Sites 15 and 17 are outliers, so removed from the analysis.

III.E.1. Fixed variables

The B31 kernel ($p = 0.046$) is significantly different from the reference B35 kernel. The standard kernel ($p = 0.15$) and the B kernel ($p = 0.86$) were not significantly different from the reference B35 kernel. There was no statistically significant difference between the 200 mAs and the 50 mAs tube current ($p = 0.75$) (Table II). There was no significant difference between the individual CT scanners ($p = 0.13$) and the baseline scanner. Mean HU values of the B31 kernel (122.2 HU) and standard kernel (121.2 HU) were higher than the B35 kernel (118.4 HU) and the B (118.8 HU) kernel. Standard errors of the means values were greater for the B35 kernel (± 1.61) than compared to the B31 (± 0.94), standard (± 0.97) and B (± 1.01) kernels (Table III).

III.E.2. Random variables

There were significant differences between individual *Test Objects* ($p < 0.0001$), CT models ($p < 0.007$), and time consistencies ($p < 0.0001$) of the individual CT scanner measurements. Standard deviation between the individual *Test Objects* (0.0004) demonstrated small differences. Differences within a given model were higher ($SD = 3.53$) than across all the scanners ($SD = 1.28$) for the water measurements. Standard deviation of the time consistency measurement (0.004) demonstrated slight variability (Table IV).

III.F. Temporal consistency

Independent of the statistical model assessment of outliers, all scanners were tracked graphically over time, allowing for the detection of temporal spikes. As shown in Figs. 5(a): internal air and 5(b): water, several sites were found to have such step changes in either air or water values. Of note is the fact that temporal anomalies in air and water values do not necessarily parallel each other. Note that site 12 shows an offset in water and then an offset in air values as the water value is brought back to an appropriate baseline. Site 37 shows a stable water value as air is seen to be offset from an appropriate baseline until it is brought back to an acceptable level in

June 2010. The sites without temporal spikes are shown in Figs. 5(c) and 5(d) representing internal air and water, respectively. Scanner manufacturers are color-coded. It is of note that while water for all scanners fell within 5 HU of the expected 0 HU value, the values of air deviated by 10 HU from the expected -1000 HU. For a scanner to be considered within tolerance at the baseline scan as criteria for entry into a study, one must consider the scanner in the context of other scanners of similar make and model.

III.G. COPDGene human subject data

As discussed above, the temporal consistency of *Test Object* scans performed on scanner 12 and 37 displayed significant changes for internal air values [Fig. 5(a)]. Thus, study subject scans from sites 12 and 37 were selected for time periods on either side of the HU shifts to assess presence of shifts in tracheal air. While it is well recognized that tracheal air varies based upon subject specific anatomy, shifts in the clustering of tracheal air values were found to coincide with shifts in the *Test Object*-based measures. An example of this is shown for scanner 12 in Fig. 6. The median and standard deviation ($\pm SD$) *Test Object* internal air attenuation measurements increased from $-1009 \text{ HU} \pm 0.3$ to $-999 \text{ HU} \pm 0.0$ after adjustment (relative to the *Test Object* measures), and mean tracheal air attenuation increased from $-984.1 \text{ HU} \pm 5.6$ to $-971.9 \text{ HU} \pm 8.7$ ($p < 0.0001$) after the shift. It is of note that when internal air measures return close to -1000 in the *Test Object*, tracheal air actually moves further away from -1000 HU.

IV. DISCUSSION

By use of a purpose developed COPDGene *Test Object*, we demonstrate significant effects of reconstruction kernels, tube current, individual *Test Objects*, CT scanner models, and temporal consistency on CT attenuation. The statistical model allows for the separation of these effects which in turn allows for an improved understanding regarding how to control and minimize these effects in future study designs as well as future scanner designs. The best way to minimize these differences is to have consistency between scanners and reconstruction kernels at the outset. The data presented here can serve as a starting point for such a discussion.

The use of fitting a linear time consistency for detecting nonstationarity in the data is a simple and powerful approach. However, it may be less sensitive in detecting subtle nonstationarity including some form of nonlinear time consistencies, short-lived level shift, etc. as demonstrated in Figs. 5(a) and 5(b). Further work on more robust detection methods is needed. Sharp calibration shifts occurring at the time of improper recalibration demonstrate that standard methods of scanner calibrations were clearly inadequate, and the sharp step-like changes show that air and water values do not necessarily track together.

Data presented here leads to the question regarding how to retrospectively correct for shifts in air and water, one logically turns to sampling of a known region of pure air or water

FIG. 5. (a) Internal air values tracked over time. (b) water tracked over time. Site 2 and 15 are Philips Brilliance 64 scanners and site 12 and 37 are GE Light Speed 16 scanners. The sites shown in these two graphs demonstrated temporal anomalies either in one or both of these two metrics. Of note is the fact that temporal anomalies in air and water values do not necessarily parallel each other. Note that site 12 shows an offset in water and then an offset in air values as the water value is brought back to an appropriate baseline. Site 37 shows a stable water value as air is seen to be offset from an appropriate baseline until it is brought back to an acceptable level in June of 2010. (c) and (d) demonstrate air and water values for the other scanners tracked in this study. While these other scanners demonstrate temporal consistency, the baseline values of water is within 5 HU of the expected 0 HU while air was within 10 HU of the expected -1000 HU.

(blood) within the reconstructed FOV. However, as shown in Fig. 4, external and internal air do not necessarily track together and, as depicted by the data shown in Fig. 6, air in the tracheal lumen can be considerably less negative than the air

FIG. 6. Tracheal air vs phantom inner air results: A shift in air attenuation (HU) demonstrated in both the human scans and phantom.

in the *Test Object*. In Fig. 6, after the scanner recalibration, tracheal air moved back to the value found at other sites that, ironically, was further away from -1000 HU. It remains unclear as to what measures of air and water (blood) should be used in a retrospective attempt to correct for scanner shifts, and these data demonstrate that the *Test Object*, while serving to characterize and monitor differences between scanners and changes over time, the *Test Object* does not fully reflect what is happening within the *in vivo* thorax.

The need for quality control in chest imaging with projection radiography was recognized some time ago and results from a *Test Object* for projection radiography was reported in 1986.¹⁴ More recently, the American College of Radiology has developed a CT *Test Object* as part of the ACR CT accreditation program.¹⁵ This *Test Object* is designed to evaluate a number of image quality parameters including positioning, CT attenuation accuracy, slice width, low contrast resolution, high contrast resolution, CT number uniformity, and image noise. This *Test Object* is used with prescribed CT protocols to assess CT examinations techniques that are used for examining the adult head, high resolution chest CT, adult abdomen, and the pediatric abdomen.

The use of a *Test Object* for correcting CT scans that are obtained for radiotherapy planning has also been reported.¹⁶ The *Test Object* is used to correct voxel by voxel the CT attenuation measurements so that they accurately reflect the electron density of the tissue imaged and then these corrected CT measurements that accurately reflect electron density of the individual tissue can be used to compute the appropriate correction factors used in radiotherapy treatments for various malignant tumors.

The use of CT attenuation measurements (HU) to quantitatively assess lung density has required a careful look at the factors involved in CT scanning protocols of the thorax that influence the lung density measurements. Using a number of different polyethylene foam densities (37–186 Kg/m³), Kemerink *et al.* showed that density resolution for both foams and lung are strongly dependent on sample volume and reconstruction kernels.^{17,18} Thicker slices, greater than 2 mm, and softer reconstruction kernels were better than thinner slices, 1 mm, and sharper reconstruction kernels in determining accurate lung equivalent foam density and lung density from CT attenuation measurements. The results of our study showed significant differences in the CT density measurements of lung foam between the different CT kernels that were used. However, it is clear that in evaluating uniform foam with small porosity, thicker slices will reduce the image noise and lead to improved density assessment. However, this is not the case for lung, especially for lung with emphysema where there is considerable variability in structure. Altering slice thickness leads to alterations in partial volume contribution to voxel densities. We thus chose to limit our test object assessment to the protocol used in scanning the COPDGene subjects. In Table II we demonstrate that, while mean differences across all scanners associated with a change in mAs is minimal, for a given scanner such protocol differences (50 vs 200 mAs) can result in as much as a 10 HU change in water and 6 HU in the lung equivalent foam. Stolk *et al.*¹⁹ have demonstrated the yearly changes in lung HU density distributions can shift by as much as 0.5–2.0 HU based upon progression of emphysema. Thus, across a multicenter study, where multiple scanners are employed, one must monitor scanning to assure not only scanner calibration, but also to assure that a given subject is scanned on the same scanner over time. If scanners are replaced over time, the change in scanner characteristics must be documented.

Our study concurred with previous findings that there are differences in comparable reconstruction kernels between different CT manufacturers and within manufacturers. Our study also showed additional random effects involving different CT models, confirming the earlier results of Stoel *et al.*²⁰ Similar to previously reported test object-based observations,²¹ our study confirmed that there are significant differences in the lung equivalent foam density using different CT scanners. Using a statistical model-based approach, we demonstrate that these effects are due to differences in reconstruction kernel, tube current as well as CT scanner model. We also confirmed that the time consistency data of the *Test Object* measurements can be used to identify abnormally operating CT scanners, four CT scanners in our study, and these CT

scanners can then be fixed and reliable results obtained once again.

The time consistency data for the different *Test Object* materials obtained on each CT scanner can be used to correct CT data and make it more robust. Gietema *et al.*²² developed a *Test Object* consisting of a 320 mm cylinder of foam with two 80 mm holes filled with air or plastic and used the results of the *Test Object* to retrospectively adjust density values. This is an area of further investigation in the COPDGene study. In this study, there was poor compliance at some imaging centers regarding regular scanning of the *Test Object*. This clearly limits the ability to use such a *Test Object* for quality assurance, and stricter measures need to be put in place to assure site compliance. With the experience presented here, the implementation of such stricter measures can now be better justified.

In conclusion, the COPDGene reference *Test Object* was able to detect a number of significant differences in the CT attenuation measurements of the *Test Object* materials as a function of the different characteristics of the 42 different CT scanners used in the COPDGene study. This enabled identification of the differences in CT scanner initial material measurements and in detecting anomalous measurement over time due to scanner malfunction. Without the concomitant imaging of *Test Objects* throughout the duration of a multicenter trial, one cannot appropriately mix data between sites, scanner make, scanner model, and reconstruction kernel type. Because of the variability across all of these parameters, it is not possible, currently, to stipulate a single HU value along with a single SD in evaluating a scan of a *Test Object* for the purposes of qualifying a particular scanner. One must take into account the scanner make, model, and kernel when qualifying a particular scanner. Once qualified, the scanner becomes its own reference to monitor temporal consistency. Statistical models such as the one introduced in this paper can be employed to help account for differences in make, model, and reconstruction kernel. In this study, we found that on a particular scanner type there were instances of temporal instability of the HU measures. It is also important to note that Figs. 5(a) and 5(b) demonstrated that site specific shifts in air and water HU metrics within the *Test Object* (for a given scanner) did not necessarily track together, suggesting that calibration shifts are not linear. This observation has important implications regarding the ability to correct the image HUs and may suggest the importance of accounting for scanner differences through statistical modeling. Because of the temporal variability, continuous monitoring, via a standardized *Test Object*, must be accounted for in the statistical model used to provide consistency across sites.

ACKNOWLEDGMENTS

The authors would like to acknowledge the technical support of Juerg Tschirren, Ph.D., Director of Engineering at VIDA Diagnostics Inc., Iowa City, IA. This study was supported, in part, by NIH-RO1-HL089897 and HL-064368.

APPENDIX: DETAILS OF THE COVARIANCE STRUCTURE

Let Y denote all N observations of the lung foam equivalent. (The covariance structure for other responses are similar.) Let B be the vector of all random effects, namely, the $m_1 = 19b_i$ s, the $m_2 b_{i,j}$ s, the m_3 pairs $(b_{i,j,k,0}, b_{i,j,k,1})$ s, in this order, always with i varying first, then j varying next, and k varying last in ordering the random effects. Let the dimensional of B be $m = m_1 + m_2 + m_3$, and B is assumed to be multivariate normally distributed with zero mean vector and a block diagonal covariance matrix, denoted by Λ , whose first m_1 diagonal elements are identically $\sigma_{\text{test object}}^2$, the next m_2 diagonal elements are identically σ_{model}^2 , and each of the next m_3 two by two symmetric matrices whose diagonal elements equal σ_0^2, σ_1^2 . Let Z be the $n \times m$ design matrix corresponding to the dummy variables of the m random effects, e.g., the first column is the dummy variable of b_1 which equals 1 for all data cases from *Test Object* 1 and 0 otherwise. Then the covariance matrix of the response vector Y equals the matrix sum $Z^T \Lambda Z + \sigma_{\text{error}}^2 I$, where I is an $n \times n$ identity matrix. The preceding covariance matrix reflects the hierarchical structure of the data, namely, *Test Object* variations, the model variation within a test subject, and scanner temporal variation within a combination of model and *Test Object*. In general, it is desirable to perform sensitivity analysis by varying the covariance structure of the model. For example, it is of interest to assess whether the reconstruction kernel or the tube current may result in difference in variance, engendering a different covariance structure. However, the sampling structure of the data is very irregular, as the *Test Objects* were scanned over irregular time intervals with different *Test Objects* scanned by different subsets of scanners, making it hard to vary the covariance structure. Indeed, we have fitted some models with more complex covariance structure that relates to the reconstruction kernel without success due to convergence problem of the estimation algorithm. This is a limitation of the data owing to the unbalanced sampling design.

Eric Hoffman is a founder and share holder, Junfeng Guo is a share holder and both John Newell and Kung-Sik Chan are consultants for VIDA Diagnostic, a company commercializing lung image analysis software developed in part at the University of Iowa.

^{a)} Author to whom correspondence should be addressed. Electronic mail: eric-hoffman@uiowa.edu

¹ D. M. Mannino, D. M. Homa, L. J. Akinbami, E. S. Ford, and S. C. Redd, "MMWR surveillance summaries: Morbidity and mortality weekly report," *MMWR CDC Surveill. Summ.* **51**(6), 1–16 (2002).

² D. L. Hoyert, E. Arias, B. L. Smith, S. L. Murphy, and K. D. Kochanek, "National vital statistics reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics," *Natl. Vital Stat. Syst.* **49**(8), 1–113 (2001).

³ D. M. Mannino and A. S. Buist, "Global burden of COPD: Risk factors, prevalence, and future trends," *Lancet* **370**(9589), 765–773 (2007).

⁴ S. S. Salvi and P. J. Barnes, "Chronic obstructive pulmonary disease in non-smokers," *Lancet* **374**(9691), 733–743 (2009).

- ⁵ A. Lokke, P. Lange, H. Scharling, P. Fabricius, and J. Vestbo, "Developing COPD: A 25 year follow up study of the general population," *Thorax* **61**(11), 935–939 (2006).
- ⁶ S. K. Alford, E. J. van Beek, G. McLennan, and E. A. Hoffman, "Heterogeneity of pulmonary perfusion as a mechanistic image-based phenotype in emphysema susceptible smokers," *Proc. Natl. Acad. Sci. U.S.A.* **107**(16), 7485–7490 (2010).
- ⁷ A. L. Friedlander, D. Lynch, L. A. Dyar, and R. P. Bowler, "Phenotypes of chronic obstructive pulmonary disease," *COPD* **4**(4), 355–384 (2007).
- ⁸ E. A. Hoffman, B. A. Simon, and G. McLennan, "State of the art. A Structural and functional assessment of the lung via multidetector-row computed tomography: Phenotyping chronic obstructive pulmonary disease," *Proc. Am. Thorac. Soc.* **3**(6), 519–532 (2006).
- ⁹ B. Patel, B. Make, H. O. Coxson, N. L. Muller, S. Pillai, W. Anderson, E. Silverman, and D. Lomas, "Airway and parenchymal disease in chronic obstructive pulmonary disease are distinct phenotypes," *Proc. Am. Thorac. Soc.* **3**(6), 533–534 (2006).
- ¹⁰ F. C. Sciurba, F. J. Martinez, R. M. Rogers, B. J. Make, G. J. Criner, R. M. Cherniack, S. A. Patel, F. Chu, H. O. Coxson, A. Sharafkhaneh, M. Elliot, J. D. Luketich, and J. C. Hogg, "Relationship between pathologic characteristics of peripheral airways and outcome after lung volume reduction surgery in severe chronic obstructive pulmonary disease," *Proc. Am. Thorac. Soc.* **3**(6), 533–534 (2006).
- ¹¹ E. A. Regan, J. E. Hokanson, J. R. Murphy, B. Make, D. A. Lynch, T. H. Beaty, D. Curran-Everett, E. K. Silverman, and J. D. Crapo, "Genetic epidemiology of COPD (COPDGene) study design," *COPD* **7**(1), 32–43 (2010).
- ¹² J. Guo, M. K. Fuld, S. K. Alford, J. M. Reinhardt, and E. A. Hoffman, "Pulmonary Analysis Software Suite 9.0: Integrating quantitative measures of function with structural analyses," in *First International Workshop on Pulmonary Image Analysis*, edited by M. Brown, M. de Bruijne, B. van Ginneken, A. Kiraly, J.-M. Kuhnigk, C. Lorenz, K. Mori, and J. M. Reinhardt (New York, 2008), pp. 283–292 (available URL: <http://www.lungworkshop.org/2009/proceedings-2008.html>).
- ¹³ J. Pinheiro and D. Bates, *Statistics and Computing* (Springer, New York, 2000).
- ¹⁴ C. Constantinou, J. Cameron, L. DeWerd, and M. Liss, "Development of radiographic chest phantoms," *Med. Phys.* **13**(6), 917–921 (1986).
- ¹⁵ C. H. McCollough, M. R. Bruesewitz, M. F. McNitt-Gray, K. Bush, T. Ruckdeschel, J. T. Payne, J. A. Brink, and R. K. Zeman, "The phantom portion of the American College of Radiology (ACR) computed tomography (CT) accreditation program: Practical tips, artifact examples, and pitfalls to avoid," *Med. Phys.* **31**(9), 2423–2442 (2004).
- ¹⁶ C. Constantinou, J. C. Harrington, and L. A. DeWerd, "An electron density calibration phantom for CT-based treatment planning computers," *Med. Phys.* **19**(2), 325–327 (1992).
- ¹⁷ G. J. Kemerink, H. H. Kruize, and R. J. Lamers, "The CT's sample volume as an approximate, instrumental measure for density resolution in densitometry of the lung," *Med. Phys.* **24**(10), 1615–1620 (1997).
- ¹⁸ G. J. Kemerink, H. H. Kruize, R. J. Lamers, and J. M. van Engelsehoven, "Density resolution in quantitative computed tomography of foam and lung," *Med. Phys.* **23**(10), 1697–1708 (1996).
- ¹⁹ J. Stolk, H. Putter, E. M. Bakker, S. B. Shaker, D. G. Parr, E. Piitulainen, E. W. Russi, E. Grebski, A. Dirksen, R. A. Stockley, J. H. Reiber, and B. C. Stoel, "Progression parameters for emphysema: A clinical investigation," *Respir. Med.* **101**(9), 1924–1930 (2007).
- ²⁰ B. C. Stoel, M. E. Bakker, J. Stolk, A. Dirksen, R. A. Stockley, E. Piitulainen, E. W. Russi, and J. H. Reiber, "Comparison of the sensitivities of 5 different computed tomography scanners for the assessment of the progression of pulmonary emphysema: A phantom study," *Invest. Radiol.* **39**(1), 1–7 (2004).
- ²¹ B. C. Stoel, F. Bode, A. Rames, S. Soliman, J. H. Reiber, and J. Stolk, "Quality control in longitudinal studies with computed tomographic densitometry of the lungs," *Proc. Am. Thorac. Soc.* **5**(9), 929–933 (2008).
- ²² H. A. Gietema, A. M. Schilham, B. van Ginneken, R. J. van Klaveren, J. W. Lammers, and M. Prokop, "Monitoring of smoking-induced emphysema with CT in a lung cancer screening setting: Detection of real increase in extent of emphysema," *Radiology* **244**(3), 890–897 (2007).