

Quality Control in Longitudinal Studies with Computed Tomographic Densitometry of the Lungs

Berend C. Stoel¹, Fred Bode², Alexis Rames³, Sherif Soliman³, Johan H. C. Reiber¹, and Jan Stolk⁴

¹Division of Image Processing, Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands; ²Medical Science, Roche Pharmaceuticals, Nutley, New Jersey; ³Roche Pharmaceuticals, Basel, Switzerland; and ⁴Department of Pulmonology, Leiden University Medical Center, Leiden, The Netherlands

To guarantee the reliability of densitometric data in clinical trials on pulmonary emphysema a quality control procedure is presented, to prevent that a measured progression in lung density could be reflected by a gradual change in the maintenance of the computed tomographic scanner. For that purpose, a foam phantom has been developed, which mimicks the densities of emphysematous lung tissue, fixed in a sealed Perspex box. Analysis software was developed to automatically compare the density readings with a baseline reference. It was found that this quality control procedure can pick up subtle changes in the scanner of less than 1 Hounsfield unit, due to changes in the X-ray tube, detectors, or reconstruction software, and can detect certain imaging artifacts. Therefore, it is recommended that this type of procedure be used to ensure the integrity of densitometric data in longitudinal studies.

Keywords: pulmonary emphysema; computed tomography; densitometry; quality control

To quantify the extent of pulmonary emphysema, densitometry using computed tomography (CT) has been applied in various studies. Through validation studies, it has been recognized that CT densitometry is influenced by parameter setting, established in the image acquisition protocol. The X-ray collimation, beam pitch, slice thickness, and reconstruction filter influence the resolution of the reconstructed images and subsequently the amount of averaging within a pixel. Because the higher the resolution of the image, the higher the noise level and the less sensitive it is to subtle density differences (1), the following general rule applies: one can either determine accurately the location of a certain structure, without knowing its density precisely, or one can determine the density accurately, without knowing its exact location because of image blurring and partial volume effects. Other influences on densitometry are based on the quality of the X-rays: a low peak kilovoltage (kVp) setting gives higher contrast, at the expense of higher noise levels. To minimize the radiation dosage of densitometric scans, the milliampere (mA) setting is typically kept low, with a risk of “photon starvation” at the apex of the lungs because of high absorption rates in the shoulders, resulting in increased noise levels.

Although information about how to select the various settings to optimize the acquisition protocol for densitometry is now available in the literature (2), no data are available concerning the stability of these settings and whether the calibration of the CT scanner is sufficient to perform accurate

follow-up studies on pulmonary emphysema. Regular water and air calibration cannot sufficiently guarantee constancy, as rigorous calibration procedures cannot prevent inconsistencies. For example, measurements of blood density in the descending aorta over time in a Dutch–Danish study (3) showed a sudden increase, not related to the disease. This was probably caused by aging of the X-ray tube (4). In another study (5), a distinct shift took place during a follow-up study, despite routine calibration and maintenance of the scanner. Even more dramatic changes in measured densities can occur when the scanner software is updated with new reconstruction filters. Even worse is when a company is taken over, whereupon all existing software is replaced by the new company’s versions (6).

In clinical drug evaluation trials with many different hospitals involved, these changes can occur frequently and will affect any density-based emphysema parameter. Therefore ideally these changes should be detected before patients are scanned, so that precautions can be taken.

This article describes a quality control procedure to detect any possible instability of a CT scanner during the course of such a study, so as to ascertain the integrity of the densitometric data. Here, we present our first experience on quality control in three clinical studies.

METHODS

We have collected quality control data from three studies: the SPREAD (Software Performance and Reproducibility in Emphysema Assessment: Demonstration) project (6) and the ongoing REPAIR (Retinoids in Emphysema Patients in the α_1 -Antitrypsin International Registry) and TESRA (Treatment of Emphysema with a γ -Selective Retinoid Agonist) clinical trials from Roche Pharmaceuticals (Basel, Switzerland). The SPREAD project, funded by the European Union, involves five centers, with one 1-detector row CT scanner and four 4-detector row scanners, all produced by the four main CT manufacturers (General Electric [Fairfield, CT], Philips [Amsterdam, The Netherlands], Siemens [Berlin, Germany], and Toshiba [Tokyo, Japan]) (7). The REPAIR study is an ongoing phase II placebo-controlled trial to investigate the efficacy, safety, and tolerability of R667 (8) in patients with symptomatic emphysema secondary to α_1 -antitrypsin deficiency. It involves 21 multi-detector row (4 or more) CT scanners, with the Philips scanner excluded because of its cutoff at $-1,000$ Hounsfield units (HU) in the density histogram, making it impossible to perform adjustments for small drifting effects in the scanner. In the TESRA trial, Philips scanners were included, because the software had been updated in the meantime. The two-arm TESRA study is investigating the efficacy, safety, and tolerability of R667 versus placebo in ex-smokers with moderate or severe emphysema. In this ongoing study 60 CT scanners are being monitored.

For all scanners in these studies, the image acquisition protocol is standardized, based on the protocol developed in the SPREAD project (9), with a radiation dose of approximately 1 millisievert (mSv) per scan. Typically, high voltage (120 kVp), low amperage (30 mAs), and a 5-mm slice thickness (increment of 2.5 mm) with a smooth reconstruction filter are used, and scanners are normally calibrated daily for air, and every 3 months for water calibration, according to the manufacturer’s guidelines.

(Received in original form April 22, 2008; accepted in final form June 10, 2008)

Supported by the Fifth Framework Programme of the European Commission (QLG1-2000-01752; the SPREAD project) and by Roche Pharmaceuticals (the REPAIR and TESRA clinical trials). The data presented in this study were obtained from these three research projects.

Correspondence and requests for reprints should be addressed to Berend C. Stoel, Ph.D., Leiden University Medical Center, Albinusdreef 2, 2333 AA Leiden, The Netherlands. E-mail: b.c.stoel@lumc.nl

Proc Am Thorac Soc Vol 5, pp 929–933, 2008
DOI: 10.1513/pats.200804-039QC
Internet address: www.atsjournals.org

A phantom was developed during the SPREAD project: a sealed Perspex box, mimicking the X-ray absorption of the thorax (Figure 1). It contains 15 compartments with pieces of polyethene foam representative of emphysematous lung tissue, ranging from 15 to 65 g/L, and is now commercially available (Medis Specials, Leiden, The Netherlands). The phantoms are scanned at each site according to the standardized protocol.

The Pulmo-CMS (Medis Specials) image analysis software package is used to analyze the resulting CT scans. First, the inner lumen of the phantom is detected fully automatically, followed by detection of the walls between the compartments; finally, the densitometric data from the slice at the center between two consecutive walls are recorded. The histogram of the densities in Hounsfield units is constructed from this slice. The ultimate density values for these pieces of foam are defined by the 5th percentile point, being the Hounsfield unit value for which 5% of all pixels have a lower density value. The 5th percentile was selected because this generated the most reproducible measurement, more so than the 15th percentile point, which is optimized for clinical rather than artificial data.

To monitor the constancy of a certain CT scanner, a baseline reference is first established, by determining the average and standard deviation in each compartment of the phantom from four CT scans, acquired over a period of 4 weeks. The standard deviation is then pooled over all compartments. This provides not only a reference for the mean values of each compartment, but also an indication of the expected stability of the scanner. If this stability cannot be guaranteed during these first 4 weeks, the CT scanner is excluded from the study, and an alternative CT scanner is sought.

During the time that a site is in the process of scanning patients for a particular study, the follow-up scans of the phantom are compared with the baseline reference. In this automated procedure, the identity of the scanner is checked together with the protocol used. If the reconstruction software is updated, a warning is generated. For each compartment the difference in density is calculated between follow-up and baseline. Subsequently, these differences are divided by the pooled standard deviation from the four baseline scans from that site. These relative differences (differences/SD) are plotted against the corresponding individual mean densities at baseline (Figure 2). If all differences remain within a range of ± 4 standard deviations, the quality of the scan is approved, because these differences are compensated for by the (*post hoc*) recalibration procedure in the image analysis software, based on the densities of blood and air, when it is applied to patient CT data (2). If the majority of the measurements are outside of this range, but still within ± 10 HU, the differences are considered significant and maintenance of the scanner is requested. The image analysis software can still compensate for this adequately, albeit with less accuracy. If the majority of the differences exceed the limits of ± 10 HU, the recalibration procedure is insufficient and the study site is asked to stop scanning patients.



Figure 1. Perspex box containing 15 compartments with pieces of foam, representative of lung tissue.

Apart from quality control of the measurements of the low-density values in the foam phantom, the CT stability for higher density values can be verified indirectly, by determining the correlation between the density values in the 15 compartments and the corresponding deviations. If a significant linear or nonlinear relation is found, it may be deduced that the deviations are increasing with increasing densities. In those cases, CT calibration for water may be needed.

The above-described procedure has been applied in the three clinical studies previously mentioned, with longitudinal data collected at each site.

RESULTS

Baseline Measurements

In Figure 3 the baseline measurements of the compartment with the highest density of foam are presented for each CT scanner

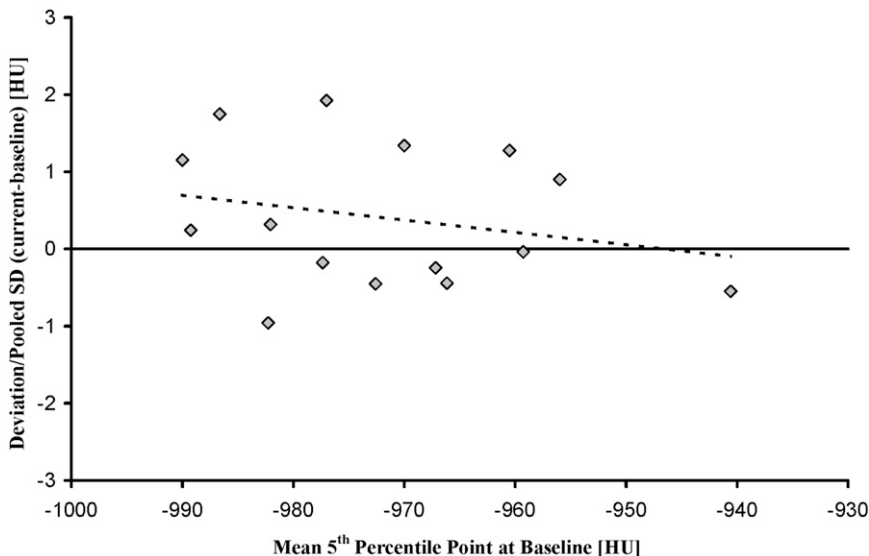


Figure 2. A graph for density quality control, in which the relative differences in density from baseline are plotted against the mean density of the four baseline scans. The dotted line indicates the regression line, which was not statistically significant in this case.

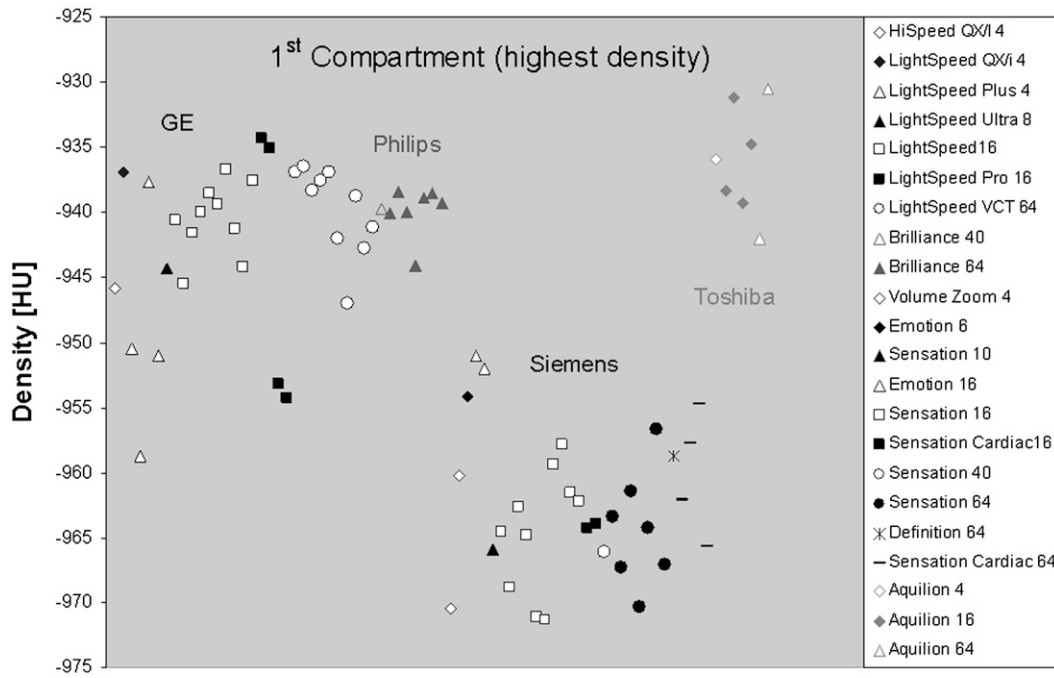


Figure 3. Measured baseline densities in the compartment with the highest foam density for each computed tomography scanner manufacturer.

used in the three studies. It clearly shows differences in measured densities between the various CT manufacturers, as was already found in an earlier study by Kemerink and coworkers (10). Siemens tends to produce consistently lower density values than the remaining companies. But also among scanners produced by one manufacturer, significant differences in density occur between the different models they have on the market. It is unlikely that these large differences are caused by differences in the production of the phantom.

Follow-up Measurements

In Figure 4, an example is given of longitudinal data from one of the majority of CT scanners that were stable over a period of more than 2 years. For each compartment the difference from the first baseline measurement was calculated and from these 15 differences the mean and its 95% confidence interval are plotted against the date of image acquisition. The three open markers give the differences of the three baseline measurements compared with the first baseline measurement.

In Figure 5, an example is given of a CT scanner that showed a gradual drift in measured densities over a period of 2 years. These changes are, however, within the safety range of ± 10 HU, for which the software program is considered to compensate effectively, when it is applied to patient CT data.

X-ray tube aging can change the density values considerably, especially at the end of its life span (Figure 6). Despite daily calibration for air, a change in density occurs just before an X-ray tube breaks down, despite daily calibration for air. The phantom analysis can, therefore, even predict when an X-ray tube is about to break down.

We also encountered some confounding factors, causing false alarms. If the table height of the CT scanner was changed, we could detect a significant difference in measured foam densities. Therefore, the table height was standardized in the image acquisition protocol for each study. Another confounder was the presence of a mounting device in the table, intended to be a head rest; it was found to influence the density readings. In some cases in which a significant correlation was found between the devia-

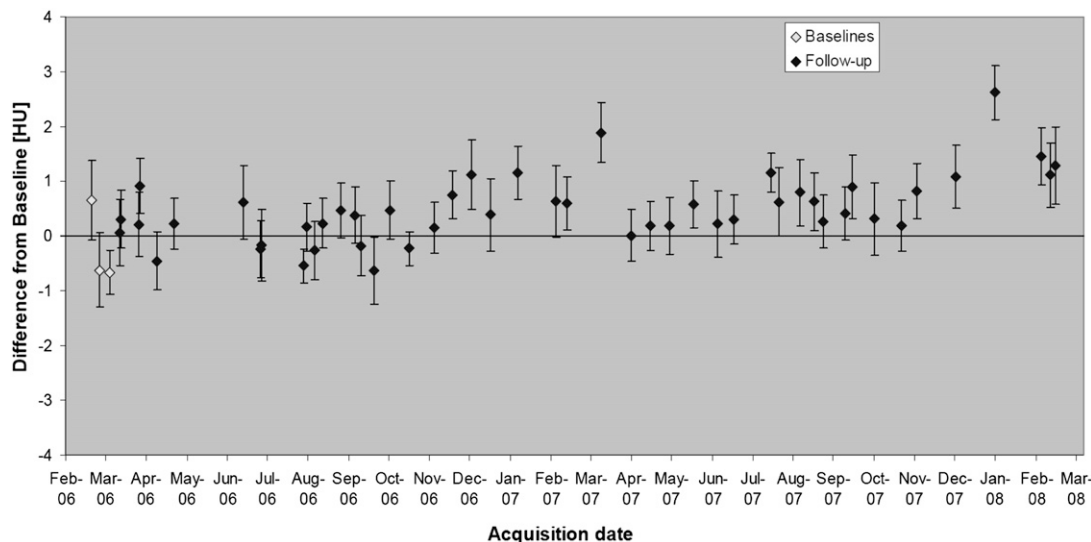


Figure 4. Example of longitudinal data from a stable computed tomography scanner over more than 2 years.

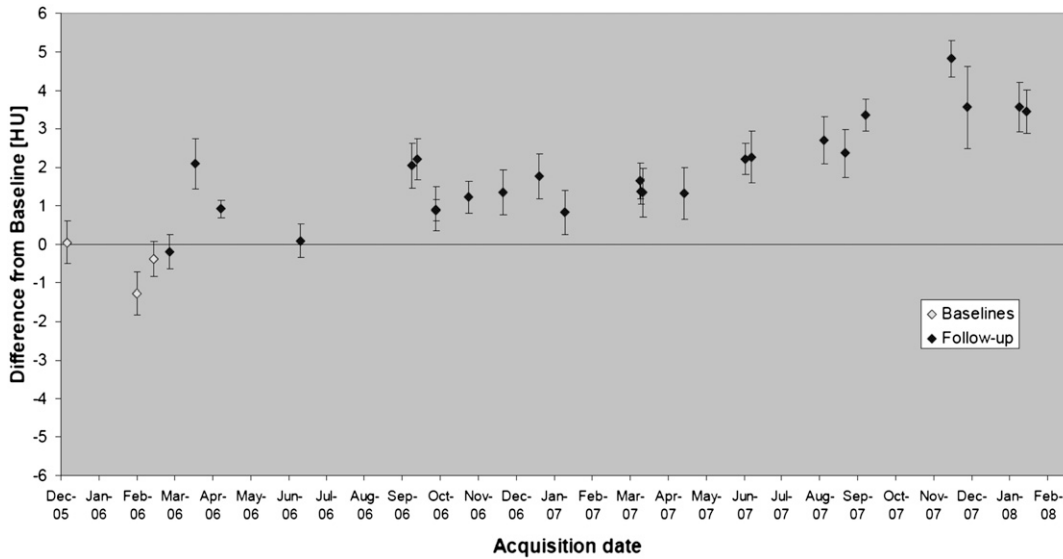


Figure 5. Example of a gradual shift in Hounsfield units over a period of 2 years.

tions from baseline and the mean densities in the compartments (suggesting insufficient water calibration) it was proven that this could be attributed entirely to the presence of these objects in the gantry.

DISCUSSION

The main purpose of CT imaging is to provide accurate visual interpretation of images based on differences in X-ray absorption between different anatomic structures, and not so much on their absolute X-ray absorption levels. To a certain extent, water and air calibration will warrant accurate X-ray absorption estimates for these extreme values. However, for intermediate densities in the range of those of lung tissue, numerous factors can influence the accuracy, despite these calibration procedures. Progression rates of 1.5 HU (6) and 2.5 HU (3) and an expected treatment effect of 1 HU/year (3) can then be easily obscured by subtle changes in a CT scanner, not accounted for in the calibration. The

American College of Radiology, for example, accepts CT scanners for certification that provide a mean density measure of air ranging from -1,005 to -970 HU (11). It is clear that these requirements for low-density values are not strict enough for the performance of clinical evaluation trials.

The quality control procedure described in this article reveals that current calibration procedures cannot prevent the CT scanner from drifting over time, mimicking disease progression. The procedure is able to track down these changes, so that precautions can be taken at an early stage. It does not yet form, however, a basis for a more elaborate calibration procedure, because the true density of the foam is difficult to determine. Variability in manufacturing the foam and compression of the pieces of foam during phantom production hinder the creation of a well-defined reference value. Therefore, the phantom was used only to check stability, and no procedure was used to standardize between CT scanners. Until now the procedure had not been used to correct for dramatic changes in a CT scanner over time, such as by re-

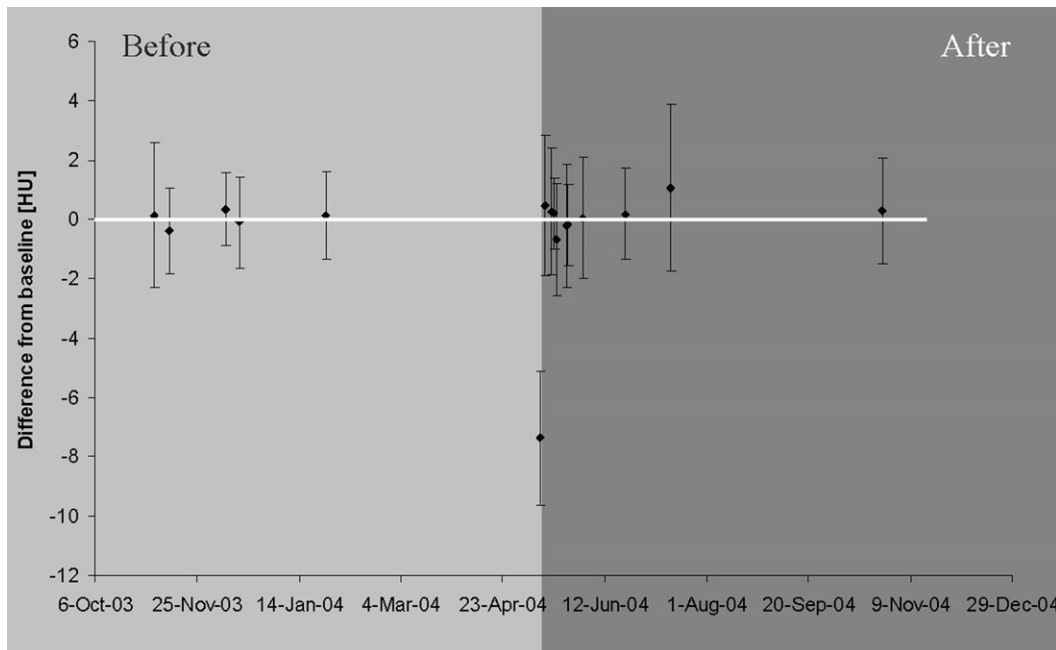


Figure 6. Sudden change in Hounsfield units, caused by an aging X-ray tube. After tube breakdown and replacement, the density values were within the normal range again.

placement with another CT scanner. It remains to be investigated whether the foam phantom can be used to correct these changes.

The stability of the phantom itself could be questioned, as polyethylene foam may degrade under the influence of light and air. Furthermore, air humidity in the foam may change over time. To control for these influences, the pieces of foam were contained in a sealed Perspex box, and the phantom itself was stored in a suitcase to be saved in a cupboard. It is, therefore, highly unlikely that the mass or volume in a compartment could change over time.

On some occasions, image artifacts, such as ringing effects and tube arcing, were found by the quality assurance procedure. These artifacts were detected automatically, because outliers in the readings caused higher variability in the measurements. However, in some cases false alarms, caused by confounding factors, could not be prevented. The fact that a mounting point in the table for the head rest can distort the phantom analysis results underlines again the importance that no other objects be present in the gantry. For example, patients should never hold their arms along the thorax, because this significantly influences the measured lung density. Furthermore, simultaneous scanning of a calibration phantom together with the patient will also influence lung densitometry and this is therefore not recommended.

The wide range of 4 standard deviations for differences was used as a safety limit, so as to prevent the too-frequent occurrence of false alarms. This consideration was made because small changes in density, albeit statistically highly significant, are already compensated for in the analysis software.

A limitation of this quality control procedure is the assumption that the baseline measurements are made with a CT scanner in a correct state of maintenance. If the CT scanner were disordered but stable during the first 4 weeks, the original state of the CT scanner could not be reconstructed after appropriate maintenance is performed. For these situations, a recalibration procedure based on the foam phantom is needed, for example, by extending the method proposed by Perhomaa and coworkers (12), by covering more than two separate density values of foam.

In conclusion, the developed phantom and accompanying image analysis procedure can guarantee the correctness of densitometric data, preventing the possibility that fluctuations in a CT scanner might lead to data suggesting alleged disease progression. Because quality control analyses are done only periodically, unreliable patient data can be labeled as such only later, after a quality control analysis has been done. However, in clinical studies, it is possible to decrease the number of unreliable CT scans by performing quality control phantom scans before obtaining patient scans. Additional research is still needed to evaluate whether the phantom can be used for calibration purposes, in case an irresolvable problem occurs.

Conflict of Interest Statement: The institution at which B.C.S. works, Leiden University Medical Center, received, in 2008, €10,000 from Bio-Imaging, €12,500 from Roche Pharmaceuticals, €59,000 from Talecris Biotherapeutics, and €14,000 from Medis Medical Imaging Systems for a research project. B.C.S. is a consultant for Roche Pharmaceuticals, Talecris Biotherapeutics, CSL Behring, and Bioimaging Technologies, Inc. F.B. is a full-time employee of Roche Pharmaceuticals. A.R. is a full-time employee of Roche Pharmaceuticals. S.S. is an employee of Hoffmann La Roche, which is a sponsor of the study. As part of his employee benefits S.S. has shares in the company that are received each year. J.H.C.R. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. J.S. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

Acknowledgments: We would like to thank all coworkers at BioImaging Technologies, Inc., Leiden, the Netherlands, involved in performing the quality control analyses for these clinical trials.

References

1. Kemerink GJ, Kruize HH, Lamers RJ, van Engelshoven JM. Density resolution in quantitative computed tomography of foam and lung. *Med Phys* 1996;23:1697–1708.
2. Stoel BC, Stolk J. Optimization and standardization of lung densitometry in the assessment of pulmonary emphysema. *Invest Radiol* 2004; 39:681–688.
3. Dirksen A, Dijkman JH, Madsen F, Stoel B, Hutchison DC, Ulrik CS, Skovgaard LT, Kok-Jensen A, Rudolphus A, Seersholm N, *et al.* A randomized clinical trial of α_1 -antitrypsin augmentation therapy. *Am J Respir Crit Care Med* 1999;160:1468–1472.
4. Stoel BC, Vrooman HA, Stolk J, Reiber JH. Sources of error in lung densitometry with CT. *Invest Radiol* 1999;34:303–309.
5. Parr DG, Stoel BC, Stolk J, Nightingale PG, Stockley RA. Influence of calibration on densitometric studies of emphysema progression using computed tomography. *Am J Respir Crit Care Med* 2004;170:883–890.
6. Stolk J, Putter H, Bakker EM, Shaker SB, Parr DG, Piitulainen E, Russi EW, Grebski E, Dirksen A, Stockley RA, *et al.* Progression parameters for emphysema: a clinical investigation. *Respir Med* 2007;101: 1924–1930.
7. Stoel BC, Bakker ME, Stolk J, Dirksen A, Stockley RA, Piitulainen E, Russi E, Reiber JHC. 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;39:1–7.
8. Belloni PN, Garvin L, Mao CP, Bailey-Healy I, Leaffer D. Effects of all-trans-retinoic acid in promoting alveolar repair. *Chest* 2000;117:235S–241S.
9. Bakker ME, Stolk J, Putter H, Shaker SB, Parr DG, Piitulainen E, Russi EW, Dirksen A, Stockley RA, Reiber JH, *et al.* Variability in densitometric assessment of pulmonary emphysema with computed tomography. *Invest Radiol* 2005;40:777–783.
10. Kemerink GJ, Lamers RJ, Thelissen GR, van Engelshoven JM. Scanner conformity in CT densitometry of the lungs. *Radiology* 1995;197: 749–752.
11. American College of Radiology. American College of Radiology CT Accreditation Program. Available from: <http://www.acr.org/accreditation/computed.aspx>
12. Perhomaa M, Jauhiainen J, Lahde S, Ojala A, Suramo I. CT lung densitometry in assessing intralobular air content: an experimental and clinical study. *Acta Radiol* 2000;41:242–248.