Inter- and intra-observer variability in the densitometric assessment of pulmonary emphysema with computed tomography

M.E. Bakker[1], B.C. Stoel[1], J. Stolk[2], H. Putter[3], A. Dirksen[4], R.A. Stockley[5], E. Piitulainen[6], E.W. Russi[7], J.H.C. Reiber[1]

[1] Division of Image Processing, Department of Radiology, [2] Department of Pulmonology, [3] Department of Medical Statistics, Leiden University Medical Center, Leiden, The Netherlands
[4] Department of Respiratory Medicine, Gentofte Hospital, Hellerup, Denmark, [5] Department of Respiratory Medicine, Queen Elizabeth Hospital, Birmingham, United Kingdom
[6] Department of Pulmonary Medicine, Malmö University Hospital, Malmö, Sweden, [7] Pulmonology Division, University Hospital Zurich, Zurich, Switzerland

Aims

Lung densitometry using Computed Tomography (CT) has proven to be an accurate measurement for assessing the progression of emphysema in clinical trials. Such studies require a high level of reproducibility in the image acquisition and in image analysis. Standardization is therefore required for the different scanners involved.
One of the aims of this study was to assess the sources of variation, originating from both acquisition and image analysis, performed in five European countries that participate in the SPREAD project.
In order to evaluate the reproducibility of densitometry, we have identified and assessed different sources of variation, i.e. inter-site, inter-scan and inter- and intra-observer variability, in comparison with the inter-subject variability. In addition we explored the influence of recalibration and volume correction on the (intra-observer) variability.

Materials & Methods

Patients and image acquisition

In five different hospitals, 119 subjects with the diagnosis of emphysema were scanned twice by CT (Table 1). Images were acquired in supine position during full inspiration, with CT settings that allow low radiation dose and high density-resolution (Stoel et al., 2004 – Invest. Radiol. 39, 1-7).

Image analysis

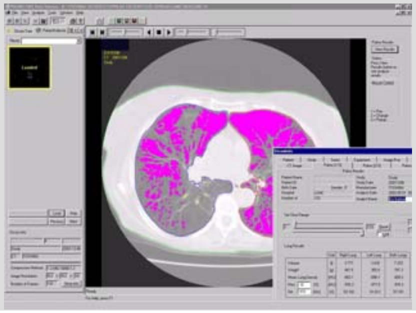
All CT images were analyzed with Pulmo-CMS (Medis medical imaging systems, Leiden, the Netherlands). At each site, an observer performed densitometry on the corresponding patient group, and in order to obtain repeated measurements for each site, a separate observer analyzed the image data from all sites. After a period of at least seven months, the coordinator reanalyzed all CT images.
In order to assess the influence of the recalibration procedure on the intra-observer variability, the coordinator analyzed the images of Site 2 with different combinations of calibration methods: no recalibration, blood recalibration only, air recalibration only, and both blood and air recalibration. These analyses were repeated after 11-18 months.

Table 1. Patient groups and scanners used in the different hospitals

Site	Patients	CT scanner
Site 1 (two observers)	24	Toshiba Aquilion 4 (MSCT)
Site 2	25	General Electric Light Speed (MSCT)
Site 3	24	Philips Mx8000 (MSCT)
Site 4	23	Philips AVEU (single slice CT)
Site 5	23	Siemens VolumeZoom (MSCT)

Statistical analysis

The density values were corrected for differences in volumes between scans according to the estimated general linear relation between log-transformed volume and Perc15 from the entire patient group per site. Variance component analysis was carried out by maximum likelihood estimation and was performed without and with volume correction. Each component is expressed as the variance with its standard error and its relative contribution to the total variance.
The intra- and inter-observer variability is presented as the mean difference and the standard deviation (SD) of the differences between observers and between different analyses from the same observer, respectively.



Results

Intra- and inter-observer variability

The values for the intra- and inter-observer variability are presented per site in Table 2.
The intra-observer variability of Sites 1 and 4 differ by as much as a factor of 150 and the inter-observer variability of these two sites differ by a factor 40.

Variance component analysis

The results of the variance component analysis are shown in Table 3. The variability between sites and between subjects together accounts for 99% of the total variation found in the densitometric data. The inter-scan variance component covered 1% of the total variation, but after volume correction the influence of this component decreases to 0.02%. The results of the variance component analysis carried out per site are presented in Table 4.

Effect of recalibration on intra-observer variability

The influence of the different recalibration methods on the intra-observer variability is presented in Table 5. Again a small but statistically significant learning effect was found when no recalibration was applied or air recalibration only (p= 0.003 and 0.002 respectively). This learning effect could not be demonstrated for the fully recalibrated data, mainly because of the increased intra-observer variability. Blood recalibration induces higher intra-observer variability than air recalibration.

Effect of air recalibration on the inter-site variability

The high inter-site variability demonstrated by the variance component analysis can be attributed mainly to the very low lung densities in the patient group of Site 1 (see Figure 1). However, this proved to be caused by a CT reconstruction problem in the Toshiba scanner, since extremely low air densities were measured outside the patient, ranging from -1060 to -1040 HU. Therefore, recalibration for air is necessary to further standardize the measurement (Figure 1). The air re-calibration reduced the inter-site variability considerably (SD=190 HU versus 137 HU).

Table 2. Intra- and inter-observer variability: mean difference and SD of the mean difference in Perc15 expressed per site

	Intra-observer		Inter-observer	
	mean difference (1st-2nd)	SD	mean difference (coord. - obs.)	SD
Site 1	0.002	0.007	0.006	0.027
Site 2	-0.001	0.207	-0.094	0.322
Site 3	-0.030	0.096	0.279*	0.388
Site 4	-0.362	1.057	-0.468*	1.067
Site 5	-0.014	0.034	-0.059*	0.099
All sites	-0.079	0.490	0.067	0.578

Table 3. Variance component analysis: variance, standard error (SE) and relative contribution of the variance components to the total variability in Perc15 (without and with volume correction).

Perc15 [HU]	No volume correction			With volume correction		
variance component	variance	SE	variance%	variance	SE	variance%
Inter-site	690.4	94.59	40.87	1046.96	93.65	47.61
Inter-subject	981.25	636.55	58.09	1151.52	743.24	52.37
Inter-scan	17.521	2.49	1.037	0.43	2.31	0.020
Inter-observer	0.011	0.01	0.0006	0.007	0.01	0.0003
Error	0.108	0.01	0.0064	0.007	0.01	0.0045

Table 4. Variance component analysis per site: relative contribution of the variance components to the total variation per site, with volume correction.

	Variance%				
Component/site	Site 1	Site 2	Site 3	Site 4	Site 5
Inter-subject	98.791	92.533	96.978	93.420	98.889
Inter-scan	1.209	7.453	3.013	6.555	1.109
Inter-observer	<< 0.001	0.003	0.003	<< 0.001	< 0.001
Error	<< 0.001	0.011	0.006	0.025	0.002

Table 5. Intra-observer variability for different recalibration procedures: mean difference and standard deviation (SD) of the mean differences in Perc15 between 2 analyses (based on dataset of Site 2).

Recalibration method	mean difference (1st - 2nd)	SD
No recalibration	-0.003*	0.004
Blood recalibration	0.005	0.141
Air recalibration	-0.018*	0.027
Both blood & air recalibration	-0.040	0.103

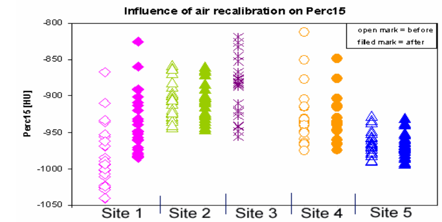


Figure 1. Effect of air recalibration on the 15th percentile point expressed per site. For site 3, air recalibration was not possible.

Conclusions

- The scan- and observer-variability is small compared to the expected progress of emphysema of 2.5 HU/year. Lung densitometry is therefore a highly reproducible measurement.
- The major effects that influence the variability of the measured density parameters originate from the site (scanner and patient group) and the subject (degree of emphysema and anatomy/physiology).
- Most reproducible lung density measurements can be obtained if:
    - the data are generated with standardized acquisition protocols and image analyses (for both longitudinal and cross-sectional studies)
    - air (and blood) recalibration is performed which further reduces the inter-site variation. without severely increasing observer variability.
    - volume correction is applied to further reduce inter-scan variation.

Acknowledgements
This study was financially supported by the Fifth Framework Programme of the European Commission (SPREAD, www.lkeb.nl/SPREAD.htm)