Robust, standardized quantification of pulmonary emphysema in Low Dose CT exams

Mario Ceresa¹, Gorka Bastarrika², Juan P. de Torres³, Luis M. Montuenga⁴, Javier J. Zulueta³, Carlos Ortiz-de-Solorzano¹, Arrate Muñoz-Barrutia¹

¹Cancer Imaging Laboratory and ⁴Biomarkers Laboratory, Oncology Division, Center for Applied Medical Research, University of Navarra, 55 Pio XII, 31008 Pamplona, Spain; ²Radiology and ³Pneumology, Clínica Universidad de Navarra, University of Navarra, 36 Pio XII, 31008 Pamplona, Spain

Telephone: +34 948 194700; Fax: +34 948 194714; E-mail address: mceresa@unav.es

Objective: Present and evaluate a fully automated system for emphysema quantification of Low Dose Computer Tomography images. The platform standardizes the emphysema measurements against changes in the reconstruction algorithm and slice thickness.

Materials and methods: Emphysema was quantified in 149 patients using a fully automatic, in-house developed software (RALPH). The accuracy of our system was evaluated against commercial software and its reproducibility was assessed using pairs of volume-corrected images taken one year apart.

Furthermore, to standardize quantifications, we modeled the effect of changing the reconstruction parameters using a non-linear fit and then applied the inverse of the model function to the data. The association between quantifications and pulmonary function tests (PFT) was also evaluated.

The accuracy of RALPH compared to the commercial software was measured using the Spearman’s rank correlation coefficient, the mean difference and the intra-subject variability. The agreement between methods was studied using the Bland and Altman plots. To assess the reproducibility of the method we used the Intra-class Correlation Coefficient (ICC) and the Bland and Altman plot. The statistical significance of the differences between the standardized data and the reference one (soft-tissue reconstruction algorithm: B40f; slice thickness: 1 mm) was assessed using a paired two-sided sampled t-test.

Results: The accuracy of our method, measured as the intra-subject variability, was 3.86 ml for pulmonary volume, 0.01 % for emphysema index (EI) and 0.39
HU for mean lung density. The reproducibility, assessed using the ICC, was higher than 0.95 for all measurements. The standardization method applied to compensate for variations in the reconstruction algorithm and slice thickness increased ICC from 0.87 to 0.97 and from 0.99 to 1.00, respectively. The correlation of the standardized measurements with PFT parameters was similar to that of the reference (for EI and the obstructive subgroup: FEV1 [%] -0.647 vs. -0.615; FEV1/FVC [%] -0.672 vs. -0.654; DLCOadj [%] -0.438 vs. -0.523).

**Conclusions:** Our new emphysema quantification method is accurate and reproducible and –thanks to our standardization method- robust against changes in the reconstruction parameters.

**Keywords:** Emphysema index, Automatic quantification, Low dose CT, COPD.

**Abbreviations:** BR, bone reconstruction; BR1, bone reconstruction algorithm and 1-mm slice thickness; BR3, bone reconstruction algorithm and 3-mm slice thickness; COPD, chronic obstructive pulmonary disease; CT, computer X-ray tomography; DLCOadj, diffusing capacity for carbon monoxide after adjusting for concentrations of hemoglobin; EI, emphysema index; EV, emphysema volume; FEV1, forced expiratory volume in the first second; FVC, forced vital capacity; FEV1/FVC, forced expiratory volume to forced vital capacity ratio; HRCT, high-resolution X-ray computer tomography; HU, Hounsfield units; CC, correlation coefficient; ICC, intra-class correlation coefficient; LDCT, low dose X-ray computer tomography; LUPA; LUng Parenchyma Analysis; MLD, mean lung density; PCC, Pearson’s correlation coefficient; PFT, pulmonary function test; RALPH, Robust Automatic on-Line Pulmonary Helper; SE, standard error; SR, soft tissue reconstruction; SR1, soft tissue reconstruction algorithm and 1-mm slice thickness; t, Kendall rank correlation coefficient; sw, inter-subject standard deviation error; TLC, total lung capacity; VOL, volume.

**Acknowledgements:** This project was partially funded by the “UTE Project CIMA”, the Spanish Ministry of Health under grants PI070751, PI070792 and RTICC RD06/0020/0066, and the subprogram for unique strategic projects of the Spanish Science and Innovation Ministry (MCIIN PSS-010000-2008-2). Mario Ceresa holds a Torres y Quevedo Fellowship of the Spanish Ministry of Science and Innovation (MICINN). Carlos Ortiz de Solorzano and Arrate Muñoz-Barrutia hold a Ramón y Cajal Fellowship of the MICINN. The authors wish to thank Carlos Jauquicoa and Xabier Artaechvarria for helpful discussions and support in technical aspects of the image processing software. They also would like to thank Elena Faedda for setting up the data transfer and implementing the anonymizer software. They are also thankful to Uxua Montes for her helpful management of the clinical data.
INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a respiratory disease with significant extra-pulmonary consequences. Its pulmonary component is characterized by airflow limitation and it is not fully reversible. The airflow limitation is associated with an abnormal inflammatory response of the lung to noxious particles or gases [1]. Traditionally, two phenotypes of COPD have been described: obstructive bronchitis and pulmonary emphysema. The former is defined as an inflammation-driven airway obstruction while the latter is characterized by parenchymal destruction. Both forms of COPD result in significant systemic co-morbidity and premature death. Due to current smoking trends and progressive ageing of the world population, an increase in COPD prevalence and related mortality is expected in the following decades [2]. Furthermore, the interest on the early detection, follow-up and distinction between those two phenotypes has been recently boosted by the discovery that COPD patients suffering emphysema have increased risk of developing lung cancer compared to those affected by airway obstruction alone [3-7].

X-ray computed tomography (CT) is commonly used to detect and measure the extent of emphysema in the lungs [8-12]. CT-based quantification of pulmonary emphysema has been shown to be reproducible; especially when the same acquisition protocol is used in all scans [10-13] and proper correction methods are employed to compensate for differences in air inspiration volumes [14]. Due to the common origins and the related incidence between emphysema and lung cancer, it seems adequate to quantify emphysema on CT images taken on ongoing lung cancer trials. Low Dose Computer Tomography (LDCT) acquisitions are part of the longitudinal protocol established for those trials in an attempt to minimize the accumulation of radiation on repeated scans.

Several automated image analysis platforms have been used to quantify emphysema in human CT scans [14-21], some of them on LDCT images [17, 18, 21]. It has been reported that CT estimates of emphysema severity depend, among other factors, on the slice thickness and reconstruction kernel used [20, 22-24]. Reconstruction protocols change with time. This is particularly relevant when
doing retrospective studies, since normally the original raw data is not kept and there is no way to redo a reconstruction using different parameters. This fact prevents a meaningful integration of emphysema measurements in cross-sectional or longitudinal studies. To the best of our knowledge, none of the automated image analysis platforms described in the literature has a mechanism to compensate for variations in the emphysema quantifications caused by changes in the reconstruction protocol.

Herein we present and evaluate a fully automated and standardized method to quantify pulmonary emphysema on CT images. The accuracy is measured as the agreement of our results with those obtained using well-established commercial software. The reproducibility of our method is evaluated by comparing same patient volume-corrected follow-up exams acquired one year apart. Furthermore, we propose a standardization method that compensates the effect of two of the most common confounding variables –reconstruction algorithm and slice thickness–.

MATERIALS AND METHODS

Subjects

The study group consisted of 149 subjects enrolled in an early lung cancer detection trial from March 2009 to November 2010. To be eligible, all subjects had to be asymptomatic, at least 40 years-old, smokers of at least 10 pack-years, with no previous history of cancer. Individuals referring clinical symptoms suggestive of lung cancer were not included in the study group. The local institutional ethics committee approved the study protocol of the lung cancer detection trial several times, the first being on 4th May 2000 and the last one on 25th February 2010. The protocol included a written informed consent from all participants.

CT acquisition

All CT exams were acquired using a 64-row multidetector CT scanner (Somatom Sensation 64; Siemens Healthcare, Forchheim, Germany) during a single breath hold at end-full-inspiration. The CT scanner was periodically calibrated using standard methods. Low-dose CT parameters were employed (120 kVp, 40 mAs)
with 2x0.6x32 mm collimation. Patients were examined in the supine position and slices were obtained contiguously from the thoracic inlet to the adrenal glands without the use of intravenous contrast material. From the raw data, non-overlapping 1-mm slice thickness images were obtained using soft tissue (SR1) and bone (BR1) reconstruction algorithms (i.e., B40f and B60f Siemens kernels). Similarly, non-overlapping 3-mm slice thickness reconstructions were performed using a bone reconstruction algorithm (BR3).

All CT images were anonymized to preserve patients’ privacy and then transferred to an external storage server (Intel Xeon, 4 core 2.5 GHz, 8 Gbytes of RAM, Linux Red Hat Enterprise Server 5.5) for off-line analysis. All the relevant information - patient’s anonymous id, study type, reconstruction kernel and slice thickness- was extracted from the DICOM image headers and stored on a database for quick reference.

**CT image analysis**

We developed the Robust Automatic on-Line Pulmonary Helper (RALPH), a software tool for the automatic, batch-mode analysis of CT images. The software proceeds in three steps:

- **Lung segmentation:** RALPH automatically extracts the internal lung volume - parenchyma and airway lumens- from the complete CT image stack using an algorithm largely based on previous work by Hu *et al.* [25]. Firstly, a threshold is iteratively computed to separate air and body tissue and remove blood vessels. Secondly, air regions connected to the borders are removed and possible holes within the lung are filled.

- **Trachea detection:** To detect the trachea, the program automatically looks for objects in the first slice of the segmented lung image. The area and the center of mass of all the objects is measured, and objects which are too big (above 400 mm²) or too small (below 100 mm²) are rejected. Then the trachea is selected from the list of remaining objects as a circular object near the center of the first slice.

- **Airway extraction:** Airway segmentation is initialized by placing a seed inside the trachea. From this seed, a wavefront propagates using a 3-D Fast Marching algorithm [26]. At each step, leakage and bifurcation checks are performed on
the propagating wavefront. Whenever a bifurcation is detected, new segments are created and propagated independently. Leakages are detected using a rule based system and immediately corrected. Finally, the centroids of the wavefront are computed to approximate the center-line of the growing segment. The process ends when all the connected voxels are visited. Then, the trachea and the main airways are deleted from the volumen. Further information on the algorithm can be found in [27]. The resulting binary image is then labeled using a 3D connected component algorithm. We will refer to this as our segmented lung image.

In parallel, a trained radiologist blindly quantified the same CT images, using the commercially available Lung Parenchyma Analysis (LUPA) image processing software (Siemens Healthcare). This program automatically detects the lung contours and the airways combining thresholding and an anatomical knowledge-based algorithm.

To measure emphysema from the segmented lung images obtained using both RALPH and LUPA, a threshold of -960 Hounsfield units (HU) was used. This threshold has been previously validated [10]. Then, the classical indexes of pulmonary emphysema quantification were obtained, including total lung volume (VOL), mean lung density (MLD) and emphysema index (EI) [28].

**Assessment of the accuracy and the reproducibility of CT-based quantification**

The accuracy of RALPH was established by measuring its agreement with LUPA, using one SR1 image per patient. The reproducibility of the emphysema quantification was assessed separately for both SR1 and BR1 by comparing the results at baseline with those obtained on a 12-month follow-up scan. Twenty-six patients out of the total 149 had no follow-up in a year’s time. Of the remaining 123 patients, 69 had the base-line scan reconstructed with SR1, whereas 54 had them reconstructed using the BR1 algorithm.

Variations in MLD and EI parameters may depend on the natural progression of the disease, and on the variability in the degree of inspiration. Therefore, before
comparison, the data was corrected for changes in the total lung volume, using a multiple linear regression, as previously described in [14]. Furthermore, the inspiratory volume was treated as a covariable with the ANCOVA formalism.

**Assessment of the sensitivity to confounding variables for CT-based quantification**

To determine the effect of the reconstruction algorithm on the quantification of emphysema, we selected one acquisition from all the patients in the study reconstructed with both SR1 and BR1 algorithms. To determine the effect of changing the slice thickness, we compared the analysis obtained on the images reconstructed using 1 mm (BR1) vs. 3 mm (BR3) slices.

The effect that the change of the reconstruction parameter had in the quantification of emphysema was modeled by least squares fitting of the scatter plot of SR1 versus BR1. The fitting parameters were selected and refined to obtain a random residual plot. Finally, we applied the inverse of the fitting function to BR1 to compensate for the effect of the reconstruction algorithm and compared the quantification results of BR1 before and after standardization with those obtained on the same image SR1 reconstructions. The same comparative analysis was performed between BR1 and BR3 groups.

**Pulmonary function test**

Forced spirometry before and after bronchodilators was performed on 65 of the 149 patients using a Vmax 22 spirometer (SensorMedics Corporation, Yorba Linda, CA, USA). The measurements were performed following the American Thoracic Society guidelines [29]. The results were expressed as a percentage of the predicted reference values given by the European Respiratory Society [30]. Patients with a postbronchodilator FEV1/FVC lower than 70% were considered to have airway obstruction following the criteria established by the Global Initiative for Chronic Obstructive Lung Disease (GOLD) [31].

**Correlation between CT parameters and pulmonary function test**

The following spirometric values were chosen for correlation with CT evaluation: force expiratory volume in 1 s (FEV1), ratio of FEV1 to forced vital capacity (FVC)
after the administration of bronchodilators, and the diffusing capacity for carbon monoxide after adjusting for the concentration of hemoglobin (DLCOadj).

**Statistical analysis**

The accuracy of RALPH compared to LUPA was measured using the Spearman’s rank correlation coefficient, which tests for the presence of a monotonical relation between the two variables. Then we performed the Bland and Altman plots, which graphically calculate the agreement and the existence of a bias between both methods using a plot of the differences versus the mean of the measurements [32]. This plot indicates whether at least 90% of the data lies within two standard deviations of the mean value. The Kendall rank correlation coefficient (τ) was applied and the inter-subject standard deviation error (sw) was calculated to test if there was any statistical dependence between the plotted variables and to estimate the confidence interval, respectively. In case of a high Kendall τ value a log-log transform was applied to the original data and the computation of sw was adapted accordingly as reported in [32].

To assess the reproducibility of RALPH measurements, we used the Intra-class Correlation Coefficient (ICC) and the Bland and Altman test. The ICC was calculated using the Cronbach’s alpha estimation method [33].

To prove that the proposed *standardization* rendered RALPH robust to a change in the reconstruction algorithm or the slice thickness, we repeated the abovementioned analysis for the comparison of two methods on SR1 vs. BR1 and BR1 vs. BR3 image groups. The statistical significance of the differences in the means of each pair of groups was assessed using a paired two-sided sampled t-test. We used the Pearson’s correlation coefficient (PCC) to correlate CT-based emphysema measurements with pulmonary function tests (PFTs).

All data analysis was performed using the open source statistical software package R [34]. A *p*-value<0.05 was considered to indicate a statistically significant difference.
RESULTS

Assessment of the accuracy and the reproducibility of the quantification

Accuracy of the quantification

Figure 1 shows the Bland and Altman plots of the quantification of VOL, MLD and EI obtained with RALPH vs. LUPA. In the EI plot we used a log-log plot to remove the dependency of the difference with the mean. As shown in the plots, most of the quantification differences lay within two standard deviations of the mean value. Table 1 shows the Spearman's coefficient, the mean difference and the intra-subject variability of VOL, MLD and EI measures for both RALPH and LUPA. There is a moderate fixed bias (i.e., mean difference) on the measurements. However, once the bias was removed, there was no statistically significant difference between measurements except for MLD, which was consistently lower when calculated by RALPH compared to LUPA (−853.65±26.65 HU vs. −835.96±29.00 HU, p=0.05).

Reproducibility of the quantification

Figure 2 (a) and (b) show the Bland and Altman plot of EI of SR1 vs. BR1, respectively. The few cases that lay outside the limits of agreement have 95% likelihood of representing a real progression of emphysema or a poorly compensated change in the inspiratory volume. Table 2 shows the ICCs being greater than 0.95 for all the measurements (VOL, MLD and EI).

Assessment of sensitivity to confounding variables for CT-based quantification of emphysema

Figure 3 shows images of representative CT slices and 3D reconstructions obtained using the same CT data set, and reconstructed with different parameters (Figure 3 (a): B40f and 1 mm; Figure 3 (b): B60f and 1 mm; (c) B60f and 3 mm). The effect that changing either the reconstruction algorithm or the slice thickness has in the EI quantification is shown in Figures 4 (a) and (b), respectively. The median EI is six times larger in BR1 compared to SR1 (5.8% vs. 28.3%) due to increased noise levels. The median EI of BR3 was lower than the
one computed for BR1 (16.8% vs. 28.6%) due to the loss of resolution in the z-axis.

**Reduction of the effect of changes in the reconstruction kernel**

Figure 5 (a) plots the EI values corresponding to SR1 versus BR1. The regression on the data was performed using a logarithmic model $Y = a \ln(x) + b$, where $a = 6.84$ and $b = 18.25$ (p-value < 0.0001). Most of the variability of the data is correctly explained by the model (see Figure 5 (b)). In consequence, a linear regression is a good fit for the standardization of the EI of BR1 and the EI of SR1 (Figure 5 (c)), being the mean difference in EIs of -0.5 % (Figure 5 (d)).

The difference between the mean of both distributions before the standardization was 22% (p-value < 0.0001). After the standardization, the difference between means descended to 0.3 % and was not significant (p-value = 0.13). The ICCs computed for the EIs before and after standardization (0.87 vs. 0.97) showed clear improvement in the agreement between the two quantifications.

The effect of changing the reconstruction algorithm resulted in a simple shift in the MLD mean of 8 HU. The ICCs before and after removing the shift were the same and equal to 0.98.

The correlation between the EI measurements and the PFTs are given in Table 3 and 4, for SR1 and BR1, respectively. The subgroup with obstructive airways disease displayed a higher correlation with the functional parameters than the control group. The same correlation for BR1 images after standardization is given in Table 5. As shown, the correlation increased after standardization of the BR1 images, achieving similar results to the correlation with SR1.

**Reduction of the effect of a change in the slice thickness**

We modeled the relationship between the EI quantifications obtained using different slice thickness (see Figure 6 (a)) using a least-squares exponential fit, i.e., $Y = a \ e^{bx} \cdot x^2$, where a=0.02 and b=-0.001 (p<0.001) (Figure 6 (b)). After the standardization, the relationship between the EIs is linear (Figure 6 (c)) and the
The width of the limits of agreement is 0.002.

The difference of means between the quantification of BR1 and BR3 (11.7%) was significant (p<0.0001). The agreement between the two quantification results was ICC= 0.99. After the slice thickness standardization, the difference of means between the two distributions remained significant (p<0.0001) but was reduced to 1.96%. The agreement had improved to ICC=1.0.

The effect of changing the slice thickness resulted in a simple shift in the mean of MLD of 2 HU. The ICCs before and after removing the shift remained unchanged and equal to 0.99.

**DISCUSSION**

Existing literature validates the use of lung densitometry for the assessment of pulmonary emphysema against histology [8, 10, 11, 28, 35] and pulmonary function [6, 11, 36-38]. A number of emphysema quantification software tools have been previously presented and extensively used. The existing commercial solutions are Pulmo CT (Siemens Healthcare, Erlangen, Germany) [10, 15, 39], the semi-automatic Volume Viewer 2 (GE Medical Systems, Milwaukee, WI, USA) [17], Pulmonary Analysis Software Suite and Emphysema Profiler (VIDA Diagnostics, Iowa City, IA, USA) [19, 23] and Pulmo-CMS (Medis, Medical Imaging Systems, Leiden, The Netherlands) [14, 36]. Several academic tools have also been described, for instance YACTA [16, 24, 37, 40], and the solutions presented by Gietema et al. [18], Boedeker et al. [20] and Keller et al. [21]. Most of them are automatic or require minimal user intervention, and rely on relatively accurate segmentation of the lungs and airways as a previous step to the emphysema quantification.

However, to the best of our knowledge, none of the existing tools incorporated a mechanism to compensate for the effects of the confounding variables that have been shown to affect the quantification results. In a first attempt to obtain a good reproducibility of the quantification, Stoel [14] suggested correcting the differences in volume of inspiration between scans and the use of the appropriate calibration protocols. However, the effects of the reconstruction
algorithm or the slice thickness changes remained unaddressed. Bodeker et al. [20] found large differences in X-ray densities of CT scans reconstructed with different algorithms. Other authors have reported the same effect for different emphysema measurements and strongly suggested using the same reconstruction set-up for the follow-up of a given patient [21, 22, 24, 39]. Gierada et al. [23] found that the quantification differences increase as a function of the magnitude of the EI. This sensitivity to changes in the reconstruction algorithm and slice thickness poses a serious problem when doing long longitudinal studies, since reconstruction protocols do change with time.

In this study, we developed an in-house system, RALPH, with the objective of obtaining a good standardization of the emphysema quantification measurements. As shown, the accuracy of RALPH was evaluated against Lung Parenchyma Analysis (Siemens Healthcare). The agreement of RALPH with respect to LUPA is good, indicating high similarity with the quantifications performed in clinical routine. The differences found were almost negligible and consisted mainly of a constant bias caused by intrinsic dissimilarities in the image processing strategies. Once the bias was removed only the MLD differences remained significant. We explain this by the fact that RALPH excluded high intensity pulmonary vessels from the lung mask when computing MLD. This seems the correct thing to do, since the mean density of parenchymal volumes can be affected by those high-density volumes. This resulted in lower RALPH MLD scores.

We have also proved that our method provides good reproducibility, as shown by the high correlation found between emphysema quantifications performed on CT images taken one year apart.

Regarding the sensitivity to changes in reconstruction algorithms and slice thicknesses, we modeled their impact on the quantification and proposed a method to minimize it. As a proof of concept of the methodology, we used the fact that in modern scanners the reconstruction parameters can be easily changed to provide several reconstructions per image acquisition. This could be easily extended to model and correct the effect of using different reconstruction parameters in retrospective cross-sectional or longitudinal studies. The
correlation between the emphysema quantifications and the PFTs increased after standardization achieving similar results to those of the reference reconstruction (kernel: B40f; slice thickness: 1 mm). Thus, our standardization method allows for a direct comparison of emphysema quantification results on images acquired on the same scanner with the same acquisition protocol but reconstructed with different parameters. Although very relevant for multi-centric studies, the effect of changing the acquisition protocol or the scanner was not addressed in this work. Repeated acquisitions on the same subject would be required, something that cannot be done on patients due to ethical concerns of X-ray overexposure. This remains as future work, and will be addressed in the future using properly designed and calibrated phantoms.

This study has several limitations. Firstly, it was a retrospective study. Our study group consisted of an open cohort of asymptomatic smokers participating in an early lung cancer detection program with low prevalence of advanced COPD. As a consequence, there are few cases with severe emphysema and thus very few data in the scatter plots have large EI values. Enrolling more patients with advanced emphysema could improve the quality of the fitting. Secondly, the reproducibility study was performed using 1-year follow-up exams, a period during which progression of emphysema might occur. However, significant progression seems unlikely [40]. Ideally, a coffee test –i.e. two scans taken minutes apart- should have been performed, but the ethical implications of non-diagnostic overexposure to X-rays prevented us from doing it. Finally, this is a single institutional study, which may limit the generalization of our findings.

In summary, we present a fully automatic and standardized method for the quantification of emphysema on LDCT images. This methodology may be useful for early detection, characterization of progression and monitoring of interventions in clinical trials. In particular, as it is fully automated and robust against changes in the reconstruction parameters, our method seems very appropriate to properly analyze large image sets in cross-sectional or longitudinal studies.
References


34. The R project for statistical computing. Web page: http://www.r-project.org/

Figure captions

Figure 1. Differences of pulmonary volume and emphysema quantification of RALPH vs. LUPA. (a) Bland and Altman plot of VOL (Bias: 136.8; Kendall t: 0.237; SE: 2.121). (b) Bland and Altman plot of MLD (Bias: 17.7; Kendall t: 0.572; SE: 0.0003). (c) Log-log Bland and Altman plot of EL (Bias: -0.209; Kendall t: 0.144; SE: 0.006). The bias is shown as a solid line; the limits of agreement are shown with a
dashed line. A value above the upper limit or below the lower limit has 95% likelihood of representing a real difference between the quantifications obtained using the two methods.

Figure 2. Differences in emphysema quantification between baseline and 12 month follow-up scan. (a) Bland and Altman plot of EI for SR1 (Bias -0.1; Kendall t: -0.105; SE 0.17). (b) Bland and Altman plot of EI for BR1 (Bias 0.6; Kendall t: -0.101; SE 0.38). The bias is shown as a solid line; the limits of agreement are shown with a dashed line. A value above the upper limit or below the lower limit of agreement has a 95% likelihood of representing a real difference between the quantification of the two exams.

Figure 3. Effect of changing the reconstruction parameters in the quantification of emphysema. Upper row: three sample transversal CT slices showing areas of density lower than -960 HU in red. Lower row: front views of the corresponding three-dimensional reconstructions, displaying main airways and lungs in transparent blue and low-density volumes in red. The reconstructions were cut with an oblique plane to show the internal distribution of the low-density volumes. All three images correspond to the same acquisition, reconstructed using different parameters (kernel and slice thickness): (left) B40f and 1 mm (VOL=13786 ml, MLD=-885 HU, EI=7.26 %); (central) B60f and 1 mm (VOL=13778 ml, MLD=-882 HU, EI=33.20 %); (right) B40f and 3 mm (VOL=1365 ml, MLD=-880 HU, EI=21.74 %).

Figure 4: Effect of confounding variables (reconstruction algorithm and slice thickness) on the quantification of the pulmonary emphysema. Box plot (minimum, 25% quartile, median, 75% quartile, maximum) of EI: (a) SR1 vs. BR1; (b) BR1 vs. BR3.

Figure 5: Standardization of the EI estimates using different reconstruction algorithms. (a) EI of SR1 vs. EI of BR1. The model that captures the relationship is drawn in solid. (b) Plot of the residuals of the fitted model. (c) EI of SR1 vs. standardized EI of BR1. (d) Bland and Altman plot of data on (c) (Bias: -0.5; Kendall t: -0.43; SE: 0.46). The bias is shown as a solid line; the limits of agreement are shown with a dashed line. A value above the upper limit or below the lower limit has a 95% likelihood represent a real difference in the quantification between the SR1 and the standardized BR1 estimates.

Figure 6: Standardization of the EI estimates using different slice thickness. (a) EI quantification on BR1 versus BR3. The exponential and linear least-squares models are drawn in solid and dashed, respectively. (b) Residual plot for the exponential fit. (c) EI of BR1 vs. standardized EI of BR3. (d) Bland Altman plot of data on (c) (Bias: 1.96; Kendall t: -0.08; SE: 0.001). Bias is shown as a solid line; the limits of agreement are shown with a dashed line. A value above the upper limit or below the lower limit has a 95% likelihood to represent a real difference in the EI quantification between the BR1 and the standardized BR3 estimates.