CHEST

Original Research

COPD

Interobserver Variability in the Determination of Upper Lobe-Predominant Emphysema*

Craig P. Hersh, MD, MPH; George R. Washko, MD; Francine L. Jacobson, MD, MPH; Ritu Gill, MBBS; Raul San Jose Estepar, PhD; John J. Reilly, MD, FCCP; and Edwin K. Silverman, MD, PhD

Background: **Appropriateness for lung volume reduction surgery is often determined based on the results of high-resolution CT (HRCT) scanning of the chest. At many centers, radiologists and pulmonary physicians both review the images, but the agreement between readers from these specialties is not known.**

Methods: **Two thoracic radiologists and three pulmonologists retrospectively reviewed the HRCT scans of 30 patients with emphysema involved in two clinical studies at our institution. Each reader assigned an emphysema severity score and assessed upper lobe predominance, using a methodology similar to that of the National Emphysema Treatment Trial. In addition, the percentage of emphysema at –910 Hounsfield units was objectively determined by density mask analysis.**

Results: **For the emphysema severity scores, (Spearman) correlation between readers ranged from 0.59 ($p = 0.0005$) to 0.87 ($p < 0.0001$), with generally stronger correlations among readers from the same medical specialty. Emphysema severity scores were significantly correlated with prebronchodilator and postbronchodilator spirometry findings, as well as with density mask analysis. In the assessment of upper lobe predominance, κ statistics for agreement ranged from 0.20 ($p = 0.4$) to 0.60 ($p = 0.0008$). Examining all possible radiologist-pulmonologist pairs, the two readers agreed in their assessments of emphysema distribution in 75% of the comparisons. Readers agreed on upper lobe-predominant disease in 9 of the 10 patients in which regional density mask analysis clearly showed upper lobe predominance.**

Conclusions: **In a group of patients with varying emphysema severity, interobserver agreement in the determination of upper lobe-predominant disease was poor. Agreement between readers tended to be better in cases with clear upper lobe predominance as determined by densitometry.**

(CHEST 2007; 131:424–431)

Key words: COPD; CT; emphysema; lung volume reduction surgery; spirometry

Abbreviations: HRCT = high-resolution CT; HU = Hounsfield units; LVRS = lung volume reduction surgery; NETT = National Emphysema Treatment Trial

Based on the results of the National Emphysema Treatment Trial (NETT)[1] and other clinical trials,[2,3] lung volume reduction surgery (LVRS) has emerged as a therapeutic option for selected patients with severe emphysema. In the NETT,[1] patients with upper lobe-predominant emphysema and a low baseline exercise capacity were found to have a survival benefit from LVRS, while those with upper lobe-predominant disease and a high baseline exercise capacity showed an improvement in symptoms and exercise tolerance following LVRS. The benefits of LVRS in subjects without upper lobe-predominant disease were marginal or nonexistent. While the high-exercise capacity and low-exercise capacity subgroups in the NETT were retrospectively defined based on the objective results of cardiopulmonary exercise testing, upper lobe-predominant emphysema was determined by an expert radiologist's interpretation of a patient's high-resolution CT (HRCT) scan of the lungs. During the NETT, the radiologist also assigned a semi-quantitative assessment of emphysema severity based on HRCT scans,[4,5] but this severity score was not included in the determination of upper lobe-predominant disease.

During a patient's evaluation for LVRS at our institution, a thoracic radiologist reads the HRCT scan of the chest, indicating the distribution and severity of emphysema, either with a categoric description (*ie*, mild, moderate, or severe) and/or with a semi-quantitative severity score. The HRCT scan is reviewed by a pulmonary physician and by a thoracic surgeon before a patient is recommended for surgery. The subjective assessment of upper lobe-predominant emphysema is a major determinant for surgical referral, but the NETT[1] and the other LVRS trials[2,3] generally relied on a single radiologist's determination of upper lobe-predominant disease. This methodology implies that emphysema distribution is an unambiguous feature of a CT scan and that multiple readers should agree in this assessment, when using a standardized grading system. In order to test this hypothesis, we investigated the agreement between multiple readers in two specialties (chest radiology and pulmonary medicine) in the description of emphysema distribution and severity on HRCT scans of patients with COPD. We then compared the readings to an objective determination of emphysema distribution based on computerized density mask analysis.

*From the Channing Laboratory (Drs. Hersh and Silverman), Pulmonary and Critical Care Division (Drs. Washko and Reilly), Thoracic Radiology Division (Drs. Jacobson and Gill), and Surgical Planning Laboratory (Dr. Estepar), Brigham and Women's Hospital, Harvard Medical School, Boston, MA.

This work was supported by US National Institutes of Health grants HL080242 (Dr. Hersh), HL71393 (Dr. Silverman), HL68926 (Dr. Silverman), and HL075478 (Dr. Silverman), a grant from the Alpha-1 Foundation (Dr. Hersh), an American Lung Association Career Investigator Award (Dr. Silverman), and a Clinical Innovator Award from the Flight Attendant Medical Research Institute (Dr. Jacobson).

Dr. Silverman has received honoraria, consultant fees, and research grants from GlaxoSmithKline for COPD genetics studies, and honoraria from Wyeth and Bayer for lectures on COPD genetics. All of the other authors have reported to the ACCP that no significant conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article.

Manuscript received April 18, 2006; revision accepted September 11, 1006.

Correspondence to: Craig P. Hersh, MD, MPH, Channing Laboratory, Brigham and Women's Hospital, 181 Longwood Ave, Boston, MA 02115; e-mail: craig.hersh@channing.harvard.edu

DOI: 10.1378/chest.06-1040

Materials and Methods

Study Subjects

CT scans and pulmonary function data were collected on 30 subjects who were enrolled in two ongoing studies of COPD at Brigham and Women's Hospital. Twenty-one subjects were from the Boston Early-Onset COPD Study; subject recruitment has been described previously.[6] Briefly, eligible probands (1) carried physicians' diagnoses of COPD, (2) were < 53 years old, (3) had an FEV_1 of < 40% predicted, (4) did not have severe α_1-antitrypsin deficiency, and (5) did not have other lung diseases. Twenty probands and one sibling from the Boston Early-Onset COPD Study were included in the study. The remaining nine subjects were enrolled in a study of chest wall physiology and obstructive lung disease in patients being evaluated for lung transplantation or LVRS.[7] In both studies, subjects provided written informed consent. Both studies were approved by the institutional review board at Brigham and Women's Hospital.

The 21 subjects in the Boston Early-Onset COPD Study completed spirometry, both prebronchodilator and postbronchodilator. The number of pack-years of smoking were derived from the study questionnaire.[8] For the other nine subjects, spirometry was performed during their clinical evaluation; postbronchodilator values were not available in all patients. Smoking history was obtained from a review of the medical record. The spirometry prediction equations of Crapo et al[9] were used for all 30 subjects.

Radiographic Analysis

Chest CT scans had been obtained in all patients for clinical indications. Examinations were performed using overlapping generations of scanners from the same manufacturer (Siemens Volume Zoom, Sensation 4, and Somatom Plus 4; Siemens Medical Solutions; Forchheim, Germany) with a full diagnostic chest CT scan protocol (*eg*, 120 to 140 kVp, typically at 237 mA, and B50 kernel reconstructions for edge enhancement). A minimum of 3 HRCT scan (1 to 2 mm) images were provided; HRCT scan images were available for many patients at 10-mm to 20-mm intervals throughout the lungs. Thick-section images (5-mm to 10-mm sections) were available in all cases. For subjects with multiple CT scans, the examination performed closest to study enrollment was used.

The 30 HRCT scan studies were independently reviewed by two thoracic radiologists (FLJ and RG) and three pulmonary physicians (CPH, GRW, and EKS). Emphysema severity was graded using a modification of the NETT scoring system,[5,10] assigning a score from 0 to 4 for the upper portion (apex to aortic arch), the mid portion (aortic arch to right inferior pulmonary vein), and lower portion (right inferior pulmonary vein to diaphragm) of each lung (see "Appendix," question A). A score of 0.5 was reserved for trivial emphysema (with < 5% of the lung affected). The craniocaudal distribution of emphysema was categorized using the NETT protocol (see "Appendix," question B). All readers were trained by one of the thoracic radiologists (FLJ), who had served as a reader in the NETT study, using separate HRCT scan studies. CT scans were interpreted using lung windows (width, –1,500 Hounsfield units [HU]; level, 600 HU) on a workstation (AGFA-Gevaert; Mortsel, Belgium). Emphysema scores were assigned primarily using the HRCT scan images; use of the thick-section images was at the discretion of the reader. Readers were aware that all patients had COPD but were blinded to other clinical characteristics.

Density mask analysis of the lung parenchyma was performed with an open-source modular software package (3D Slicer; www.Slicer.org), based on the contiguous thick-section images.

In the software, automatic extraction of the lung is achieved by performing thresholding using the Otsu method,[11] which finds the optimal threshold to separate the image into two classes by analyzing the image histogram. After that, the centroids of the lower intensity regions are used to identify the connected components and to extract the left and right lungs. Gaps in the lung mask are removed by means of morphologic operations. Finally, vessels are extracted by applying the Otsu method in the lung area and removing those pixels corresponding to the upper threshold. The resulting lung mask is divided into three regions of equal volume to enable regional density mask analysis. This method for the extraction of the lung mask is comparable to other methods reported in the literature[12]; the main difference is related to use of the Otsu method for the automatic definition of the thresholds used to separate the image into two classes. No attempt was made to exclude airways from the mask and therefore from the density mask analysis.

Density mask analysis was used to calculate the fraction of emphysematous lung on the CT scan at a threshold of –910 HU[13] for the upper, middle, and lower third of each lung. A subject was defined as having definite upper lobe-predominant emphysema if the percentage of emphysema in the upper third of the lung (averaged across both lungs) was at least 10% greater than the percentage of emphysema in both the middle and lower thirds of the lung in order to ensure a clear distinction. Non-upper lobe-predominant disease was present if the percentage of emphysema in the upper third of the lung was less than that in the middle or lower third. If the percentage of emphysema in the upper third of the lung was greater than that in both the middle and lower thirds, but not at least 10% greater, then the distribution was called *borderline*. A density mask threshold of −950 HU was used in a secondary analysis.

Statistical Analysis

Agreement between readers for emphysema scores was assessed using Spearman correlation coefficients. Emphysema distribution was classified as upper lobe-predominant or non-upper lobe-predominant, as in the NETT. Agreement between readers was assessed by κ statistics, using exact p values. Based on the nonnormal distribution of emphysema scores and quantitative image analysis results, the relationships with clinical characteristics were calculated using nonparametric statistics (*ie*, Spearman correlation, Wilcoxon rank sum test, or Kruskal-Wallis test, as appropriate). A p value of < 0.05 defined statistical significance. Analyses were performed using a statistical software package (SAS, version 9.1; SAS Institute; Cary, NC).

Results

All 30 subjects had a history of cigarette smoking. Airflow obstruction ranged from mild (prebronchodilator FEV_1, 73% predicted) to very severe (prebronchodilator FEV_1, 10% predicted) [Table 1]. Emphysema scores, averaged across five readers, ranged from 5.9 to 20.6 (possible maximum score, 24). In the 21 subjects from the Boston Early-Onset COPD Study, the mean emphysema score ranged from 6.8 to 20.6; 8 of 21 subjects had a score of < 12, signifying mild-to-moderate emphysema.

Interobserver correlations for emphysema severity scores ranged between 0.59 and 0.87 (Table 2); all correlations were statistically significant. In general, agreement was stronger among readers in the same specialty; correlation between the two thoracic radiologists was 0.76 ($p < 0.0001$), and among the three pulmonary physicians ranged from 0.82 to 0.87 (all $p < 0.0001$).

κ statistics for agreement between readers in the assessment of upper lobe-predominant emphysema ranged from 0.20 ($p = 0.4$) to 0.60 ($p = 0.0008$) [Table 3]. Agreements were slightly better among observers in the same specialty. The κ statistic for the two radiologists was 0.34 ($p = 0.1$), and among the three pulmonologists ranged from 0.44 ($p = 0.02$) to 0.59 ($p = 0.002$). All five readers agreed on emphysema distribution in 14 cases (46.7%). Four of five readers agreed in 11 cases (36.7%), and three of five readers agreed in 5 cases (16.7%) [Fig 1].

In order to simulate the clinical assessment of a patient undergoing LVRS evaluation, we examined all pairwise combinations of one radiologist's and one pulmonary physician's determinations of emphysema distribution. Of the 180 possible pairs (two radiologists and three pulmonary physicians), agreement was found in 135 pairs (75.0%).

Emphysema scores averaged across all five readers were not significantly related to sex or age (Tables 4, 5); there was a trend for correlation with number of pack-years of smoking (Spearman $r = 0.33$; $p = 0.07$). Emphysema scores were inversely correlated with FEV_1 and FEV_1/FVC ratio, although the correlations were moderate (−0.35 to –0.49). Despite the smaller sample size, the correlations were greater for postbronchodilator spirometry than for prebronchodilator spirometry.

In 25 subjects, at least four of the five readers agreed in the assessment of emphysema distribution (Table 5). Eighteen of the 25 subjects had upper

Table 1—*Characteristics of Study Subjects**

Characteristics	Values
Sex	
Female	16 (53.3)
Male	14 (46.7)
Age	52.3 ± 8.2
Smoking intensity, pack-yr	47.4 ± 27.0
FEV_1, % predicted	
Prebronchodilator	23.1 ± 11.9
Postbronchodilator†	25.4 ± 14.7
FEV_1/FVC ratio	
Prebronchodilator	0.29 ± 0.08
Postbronchodilator†	0.29 ± 0.08
Emphysema score (average of five readers)	14.0 ± 4.3
Emphysema % − 910 HU	57.1 ± 12.3

*Values are given as No. (%) or mean ± SD. n = 30, except as noted.
†n = 22 (eight subjects did not undergo postbronchodilator spirometry).

Table 2—*Correlations Between Readers for Emphysema Severity Scores**

Readers	R1	R2	P1	P2	P3
R1		0.76 (< 0.0001)	0.76 (< 0.0001)	0.84 (< 0.0001)	0.66 (< 0.0001)
R2			0.59 (0.0005)	0.65 (0.0001)	0.59 (0.0006)
P1				0.82 (< 0.0001)	0.87 (< 0.0001)
P2					0.84 (< 0.0001)

*Values are given as Spearman correlation coefficients (p values). R1 = radiologist 1; R2 = radiologist 3; P1 = pulmonary physician 1; P2 = pulmonary physician 2; P3 = pulmonary physician 3.

lobe-predominant emphysema, and 7 subjects had non-upper lobe-predominant emphysema. Average emphysema severity scores were significantly higher in the 25 subjects in whom a consensus emphysema distribution could be determined (consensus distribution, 14.9; no consensus, 9.3; p = 0.01).

In the computerized density mask analysis, the mean percentage of emphysema at −910 HU was 57.1% (Table 1), with a range from 24.0 to 73.1%. The density mask results were inversely correlated with FEV_1 and FEV_1/FVC ratio (Table 4). The strength of the correlations between these spirometric traits and densitometry (−0.36 to −0.54) was similar to the strength of the correlations of emphysema severity scores with spirometry. The percentage of emphysema at –910 HU was significantly correlated with the average semi-quantitative severity score ($r = 0.72$; $p < 0.0001$). Similar to the severity score analysis, the percentage of emphysema was higher in subjects in whom a consensus description of emphysema distribution could be reached (Table 5); automated analysis results showed a trend toward significance.

Based on the regional density mask analysis (−910 HU), 10 of the 30 CT scans showed upper lobe-predominant emphysema (Table 6). Agreement between the human consensus readings and the computerized analysis was moderate ($\kappa = 0.26$; p = 0.03). Human readers (at least four of five) agreed on upper lobe-predominant disease in 9 of 10 cases in which densitometry clearly showed upper lobe predominance, yet readers agreed on upper lobe predominance in only 8 of 14 cases in which densitometry was borderline (p = 0.17 [Fisher exact test]). Based on a less strict definition of upper lobe-predominant disease that did not require the 10% difference threshold (*ie*, combining the clearly upper lobe and borderline categories), 24 cases were called upper lobe-predominant by densitometry findings. Readers agreed on upper lobe predominance in only 17 of these 24 cases.

As a confirmatory analysis, agreement between human readers was also compared with upper lobe emphysema, as defined by a −950-HU density mask threshold. Human readers agreed in 14 of 15 cases in which densitometry showed upper lobe predominance using the −950-HU threshold. However, readers agreed in only 4 of 11 cases in which densitometry at −950 HU was borderline upper lobe-predominant (p = 0.003).

Discussion

In a review of HRCT scans in 30 COPD patients, we found good correlation among two thoracic radiologists and three pulmonary physicians in terms of semi-quantitative emphysema severity scores, but found poor agreement in the determination of upper lobe-predominant emphysema. Among the cases with clear upper lobe predominance determined by regional density mask analysis, there was better agreement among the human readers. The disagreement tended to be found in subjects with marginal upper lobe predominance (*ie*, a < 10% difference between upper and middle/lower thirds of the lung) according to the densitometry results.

Emphysema severity scores averaged across the five readers were significantly correlated with spirometric measures of airflow obstruction, with stronger correlations for postbronchodilator spirometry values. However, these correlations were only moder-

Table 3—*Agreement Between Readers in Assessing Upper Lobe-Predominant Emphysema**

Readers	R1	R2	P1	P2	P3
R1		0.34 (0.1)	0.51 (0.007)	0.33 (0.1)	0.20 (0.4)
R2			0.59 (0.002)	0.60 (0.0008)	0.43 (0.05)
P1				0.49 (0.003)	0.59 (0.002)
P2					0.44 (0.02)

*Values are given as κ statistics (p values). See Table 2 for abbreviations not used in the text.

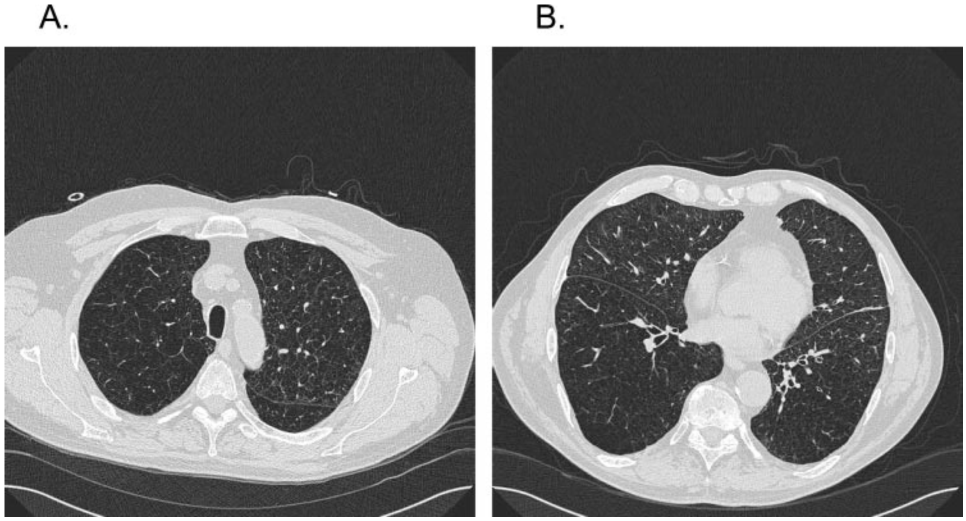

FIGURE 1. Example of disagreement among readers. Two of five observers interpreted this HRCT scan as showing upper lobe-predominant emphysema. Representative 1-mm sections from the level of (*left*, *A*) the top of the aortic arch and (*right*, *B*) the right inferior pulmonary vein are shown, using lung windows (width, –1,500 HU; level, 600 HU).

ate. Because of the poor interobserver agreement in the assessment of upper lobe-predominant disease, we did not attempt to compare pulmonary function across the different classes of emphysema location.

Previous authors have examined the agreement between multiple readers in the assessment of emphysema distribution. Slone and colleagues[14] retrospectively reviewed the CT scans of 50 patients who had undergone LVRS. Four chest radiologists graded emphysema heterogeneity and severity on a scale of 0 to 4. Correlation between readers was strong for both measurements (Pearson $r = 0.82$ for heterogeneity; $r = 0.75$ for severity). In a study by Weder et al,[15] five "clinicians" and one radiologist retrospectively evaluated preoperative CT scans from 50 LVRS patients. Emphysema distribution was defined as markedly heterogeneous, intermediately heterogeneous, or homogeneous. On average, 5.4 of the 6 readers agreed on the assessment of markedly heterogeneous disease, which included upper lobe-predominant disease. In two other studies, differences in emphysema severity scores between lung regions were used to describe heterogeneity. Wisser and coauthors[16] found interobserver κ statistics ranging from 0.54 to 0.79 among three readers, while Pompeo and colleagues[17] reported κ coefficients of 0.67 to 0.92 for pairwise comparisons between two radiologists and a surgeon.

In these published analyses, the agreement among multiple readers in different specialties was better than we found in the present study. However, the previous studies were all retrospective reviews of preoperative CT scans in patients who had undergone LVRS. These patients are likely to have more

Table 4—*Correlations (Spearman) of Emphysema Severity Scores and Density Mask Analysis With Clinical Characteristics and Pulmonary Function Test Results**

	Severity Score		Emphysema % − 910 HU	
Characteristics	Correlation Coefficient	p Value	Correlation Coefficient	p Value
Age	0.29	0.1	0.36	0.05
Pack-yr of smoking	0.33	0.07	0.09	0.7
FEV_1 % predicted				
Prebronchodilator	− 0.40	0.03	− 0.36	0.05
Postbronchodilator†	− 0.47	0.03	− 0.54	0.01
FEV_1/FVC ratio				
Prebronchodilator	− 0.35	0.06	− 0.39	0.03
Postbronchodilator†	− 0.49	0.02	− 0.51	0.01
Emphysema % − 910 HU	0.72	< 0.0001		

*n = 30, except as noted.
†n = 22 (eight subjects did not undergo postbronchodilator spirometry).

Table 5—*Mean Emphysema Severity Scores and Percentage of Emphysema (−910 HU) Across Categoric Variables*

Characteristics	Patients, No.	Severity Score Mean ± SD	Severity Score p Value	Emphysema % − 910 HU Mean ± SD	Emphysema % − 910 HU p Value
Sex			0.5*		0.06*
Female	16	13.2 ± 5.2		52.7 ± 14.4	
Male	14	14.8 ± 3.1		62.1 ± 7.1	
Emphysema distribution‡			0.03†		0.1†
Upper lobe-predominant	18	15.5 ± 3.4		59.5 ± 10.4	
Non-upper lobe-predominant	7	13.3 ± 4.9		59.1 ± 10.8	
No consensus	5	9.3 ± 3.2		45.6 ± 16.5	
Emphysema distribution‡			0.01*		0.05*
Consensus	25	14.9 ± 3.9		59.4 ± 10.3	
No consensus	5	9.3 ± 3.2		45.6 ± 16.5	

*Wilcoxon rank sum test,
†Kruskal-Wallis test.
‡Based on agreement among at least four of five readers.

severe emphysema, with upper lobe predominance, based on commonly accepted indications for LVRS prior to the completion of the NETT.[18] With less variability between patients, interobserver agreement should be improved. However, in our study, we reviewed HRCT scans of patients who did not necessarily undergo LVRS. There was more variability between patients, which is likely to explain the greater divergence in the assignment of upper lobe-predominant disease. Our patient cohort may reflect more accurately the patient population undergoing initial LVRS evaluation, at the point where the subjective classification of emphysema distribution and severity may greatly impact eventual surgical referral. There are other anatomic features assessed by HRCT scan that may also impact a patient's appropriateness for LVRS, such as the extent of destruction in the portions of the lungs that will not be resected. However, without a standardized scoring system, we were unable to assess these factors.

Despite the wider range of emphysema severity, correlation among the five readers in our study was generally good. In general, interobserver agreement, using several different methods of emphysema severity scoring, has not been as good in studies of COPD patients who were not being evaluated for LVRS (reviewed by Malinen et al[19]). This also may be due to the limited variation in emphysema severity in LVRS patients. In our study, interobserver agreement tended to be better in the patients with more severe emphysema. We also found the agreement in the assessments of both distribution and severity to be generally better among physicians from the same specialty. This may reflect different clinical experience in chest CT scan interpretation, since all readers in our study were uniformly trained with our scoring method at the same sessions. Despite the variability, the emphysema severity score, when averaged across the five readers, was strongly correlated to the results of computerized density mask analysis, using the threshold of –910 HU (defined *a priori*), which is a commonly used threshold for the quantification of emphysema.[20]

The average emphysema severity score showed moderate, but significant, correlation with measures of airflow obstruction. The correlations we found with FEV_1 and FEV_1/FVC ratio are similar to the values found by other investigators.[10,21,22] The severity of the emphysema seen on CT scans explains only a portion of the variability in spirometry results. Other factors, such as airway disease, are likely to be important. Even among the subjects from the Boston

Table 6—*Agreement Between Human and Computerized Assessment of Upper Lobe-Predominant Emphysema*

Emphysema Distribution*	Patients, No.	Computerized Density Mask Assessment† Non-Upper Lobe-Predominant	Upper Lobe-Predominant	Borderline
Non-upper lobe-predominant	7	3	1	3
Upper lobe-predominant	18	1	9	8
No consensus	5	2	0	3

*Based on agreement between at least four of five readers. Emphysema distribution was assessed based on question B in the "Appendix."
†Based on the percentage of emphysema at –910 HU.

Early-Onset COPD Study, who had severe airflow obstruction, there was still substantial variability in emphysema severity scores.

Since the CT scans in our study were performed for clinical indications over a period of several years, the CT scanner models were not the same for all 30 patients. Because of different imaging protocols, there were variable numbers of HRCT scan images for each patient. These factors may contribute to the variability we observed. However, this closely reflects the clinical evaluation for LVRS, in which patients may undergo CT scans at different institutions or on different scanners at the same institution. Different radiologists and pulmonologists may be evaluating patients at the same center, a situation that is reproduced by our study design. Another limitation of our study was the absence of postbronchodilator spirometry results for 8 of the 30 subjects. Despite the reduced sample size, the correlations with emphysema score were stronger for postbronchodilator variables. Emphysema may contribute more to the fixed aspect of airflow obstruction, reflected by the postbronchodilator spirometry results.

Several groups have proposed quantitative CT scan methods to define upper lobe-predominant emphysema, usually based on densitometry differences between the upper and lower portions of the lungs.[23–26] In two small studies,[27,28] the ratio of the percentage of emphysema (−900 HU) in the upper half of the lungs to the lower half of the lungs was found to correlate with improved FEV_1 following LVRS. In an analysis of the NETT data,[29] quantitative CT scan assessment was at least as good as the radiologist's interpretation in predicting the response to LVRS. We were able to use density mask analysis to identify a source of variability in the readers' assessments; namely, the distinction between the clear upper lobe-predominant cases and the borderline upper lobe-predominant cases.

Based on the results of the NETT[1] and other trials,[2,3] patients with upper lobe-predominant emphysema may be expected to benefit from LVRS. In the NETT, the determination of upper lobe-predominant disease was based on a single radiologist's interpretation, without accounting for the variability in HRCT scan reading that we have demonstrated in the present study. Using computerized density mask analysis as a comparison, we have found that interobserver agreement was improved in patients with clear upper lobe-predominant emphysema vs those with more marginal upper lobe predominance. One might expect that LVRS may be more beneficial in the former group of patients than in the latter group. Further analysis of data from the NETT and other LVRS trials will be required to answer this question. If this were shown to be the case, then formal review of the HRCT scan by multiple readers might be recommended before a patient is referred for LVRS.

ACKNOWLEDGMENT: The authors thank Bernard Rosner for statistical advice, and Scott Weiss, Frank Speizer, Jeffrey Drazen, Hal Chapman, Leo Ginns, and Steve Mentzer for their assistance in developing the Boston Early-Onset COPD Study.

Appendix

Grading of Emphysema Severity and Distribution on HRCT Scans

A. Emphysema Severity Scale: Scores were assigned for the upper, mid, and lower portions (see the "Materials and Methods" section for definitions) of the right and left lung.

0. Normal (none)
0.5. Trivial (< 5% of lung affected):

1. Mild (5–25%)
2. Moderate (26–50%)
3. Marked (51–75%)
4. Severe (> 75%)

B. Best Description of Craniocaudal Distribution of Emphysema:

1. Upper lobe-predominant
2. Lower lobe-predominant
3. Diffuse
4. Superior segments of lower lobes predominantly involved

References

1 Fishman A, Martinez F, Naunheim K, et al. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. N Engl J Med 2003; 348: 2059–2073
2 Miller JD, Berger RL, Malthaner RA, et al. Lung volume reduction surgery vs medical treatment: for patients with advanced emphysema. Chest 2005; 127:1166–1177
3 Berger RL, Wood KA, Cabral HJ, et al. Lung volume reduction surgery: a meta-analysis of randomized clinical trials. Treat Respir Med 2005; 4:201–209
4 National Emphysema Treatment Trial Research Group. Rationale and design of The National Emphysema Treatment Trial: a prospective randomized trial of lung volume reduction surgery. Chest 1999; 116:1750–1761
5 National Emphysema Treatment Trial Research Group. Patients at high risk of death after lung-volume-reduction surgery. N Engl J Med 2001; 345:1075–1083
6 Silverman EK, Chapman HA, Drazen JM, et al. Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease: risk to relatives for airflow obstruction and chronic bronchitis. Am J Respir Crit Care Med 1998; 157: 1770–1778
7 Ingenito EP, Evans RB, Loring SH, et al. Relation between preoperative inspiratory lung resistance and the outcome of lung-volume-reduction surgery for emphysema. N Engl J Med 1998; 338:1181–1185
8 Ferris BG. Epidemiology Standardization Project (American Thoracic Society). Am Rev Respir Dis 1978; 118:1–120
9 Crapo RO, Morris AH, Gardner RM. Reference spirometric values using techniques and equipment that meet ATS

recommendations. Am Rev Respir Dis 1981; 123:659–664
10 Bergin C, Muller N, Nichols DM, et al. The diagnosis of emphysema: a computed tomographic-pathologic correlation. Am Rev Respir Dis 1986; 133:541–546
11 Otsu N. Threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 1979; 9:62–66
12 Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans Med Imaging 2001; 20:490–498
13 Muller NL, Staples CA, Miller RR, et al. "Density mask": an objective method to quantitate emphysema using computed tomography. Chest 1988; 94:782–787
14 Slone RM, Pilgram TK, Gierada DS, et al. Lung volume reduction surgery: comparison of preoperative radiologic features and clinical outcome. Radiology 1997; 204:685–693
15 Weder W, Thurnheer R, Stammberger U, et al. Radiologic emphysema morphology is associated with outcome after surgical lung volume reduction. Ann Thorac Surg 1997; 64:313–319
16 Wisser W, Klepetko W, Kontrus M, et al. Morphologic grading of the emphysematous lung and its relation to improvement after lung volume reduction surgery. Ann Thorac Surg 1998; 65:793–799
17 Pompeo E, Sergiacomi G, Nofroni I, et al. Morphologic grading of emphysema is useful in the selection of candidates for unilateral or bilateral reduction pneumoplasty. Eur J Cardiothorac Surg 2000; 17:680–686
18 Cooper JD, Lefrak SS. Lung-reduction surgery: 5 years on. Lancet 1999; 353(suppl):SI26–SI27
19 Malinen A, Erkinjuntti-Pekkanen R, Partanen K, et al. Reproducibility of scoring emphysema by HRCT. Acta Radiol 2002; 43:54–59
20 Muller NL, Coxson H. Chronic obstructive pulmonary disease: 4. Imaging the lungs in patients with chronic obstructive pulmonary disease. Thorax 2002; 57:982–985
21 Sanders C, Nath PH, Bailey WC. Detection of emphysema with computed tomography: correlation with pulmonary function tests and chest radiography. Invest Radiol 1988; 23:262–266
22 Eda S, Kubo K, Fujimoto K, et al. The relations between expiratory chest CT using helical CT and pulmonary function tests in emphysema. Am J Respir Crit Care Med 1997; 155:1290–1294
23 Saitoh T, Koba H, Shijubo N, et al. Lobar distribution of emphysema in computed tomographic densitometric analysis. Invest Radiol 2000; 35:235–243
24 Cederlund K, Tylen U, Jorfeldt L, et al. Classification of emphysema in candidates for lung volume reduction surgery: a new objective and surgically oriented model for describing CT severity and heterogeneity. Chest 2002; 122:590–596
25 Coxson HO, Whittall KP, Nakano Y, et al. Selection of patients for lung volume reduction surgery using a power law analysis of the computed tomographic scan. Thorax 2003; 58:510–514
26 Stavngaard T, Shaker SB, Dirksen A. Quantitative assessment of emphysema distribution in smokers and patients with α_1-antitrypsin deficiency. Respir Med 2006; 100:94–100
27 Gierada DS, Slone RM, Bae KT, et al. Pulmonary emphysema: comparison of preoperative quantitative CT and physiologic index values with clinical outcome after lung-volume reduction surgery. Radiology 1997; 205:235–242
28 Flaherty KR, Kazerooni EA, Curtis JL, et al. Short-term and long-term outcomes after bilateral lung volume reduction surgery: prediction by quantitative CT. Chest 2001; 119: 1337–1346
29 Hoffman EA. Quantitative image analysis in predicting outcomes in NETT. Paper presented at: American Thoracic Society International Conference; May 21–26, 2004; Orlando, FL