CT quantification of emphysema: Is semi-quantitative scoring a reliable enough method?

Laila A. Mohsen *, Ehab A.A. Gawad, Mohamed A. Ibrahiem

Department of Diagnostic Radiology, Al-Minya University Hospital, Egypt

Received 27 July 2013; accepted 25 March 2014
Available online 22 April 2014

Key words: COPD; CT emphysema index; Modified Goddard score; Mean lung attenuation; Emphysema.

Abstract Objective: To evaluate whether a simple semi-quantitative method aided by software enhanced visualization can be reliable enough for the quantification of emphysema during the daily workload.
Patients and methods: Thirty patients with COPD were included. Patients had a standard non enhanced MDCT study of the chest using a 16 slice machine. The images were evaluated visually and scored. This scoring was repeated after applying a density mask. Three radiologists evaluated the images on separate occasions. Repeatability was also tested. The CT emphysema index and the mean lung attenuation were calculated. The extent of airway disease was not assessed.
Results: Kappa test between the 3 readers revealed slight agreement \( k = 0.122, p = 0.001 \) before the density mask and substantial agreement \( k = 0.75, p < 0.0005 \) after its application. A high degree of repeatability was found. The median visual score after density mask application, showed a stronger correlation to the emphysema index \( r = 0.81, p < 0.0005 \) than before.
Conclusion: We present a simple visual score for quantitation of emphysema, that when combined with a simple density mask, the inter-rater agreement and repeatability of scoring are markedly improved. This method appears to be fast and easy to perform.

1. Introduction

Chronic obstructive pulmonary disease is a common health problem that we encounter in everyday patients. The pathogenesis of the disease is still not fully understood and is probably a combination of several factors. The factors that were frequently under study are the degree of airway obstruction and the extent of emphysema.

The volumetric CT quantification of emphysema is not part of the routine chest CT study in COPD patients until now. The visual evaluation methods are highly dependent on the radiologist and his experience level. The quantitative methods, on the other hand, are time-lengthy and require certain image

Abbreviations: COPD, chronic obstructive pulmonary disease; MLA, mean lung attenuation; GOLD, global initiative for chronic obstructive lung disease; MDCT, multi-detector computed tomography; FEV1, forced expiratory volume in 1 second; FVC, forced vital capacity; LVRS, lung volume reduction surgery.

* Corresponding author. Address: Al-Minya University, Al-Minya, Egypt. Tel.: + 20 1064308545. E-mail address: dr.laila.adel77@gmail.com (L.A. Mohsen).

Peer review under responsibility of Egyptian Society of Radiology and Nuclear Medicine.

0378-603X © 2014 Production and hosting by Elsevier B.V. on behalf of Egyptian Society of Radiology and Nuclear Medicine.

Open access under CC BY-NC-ND license. http://dx.doi.org/10.1016/j.ejrnmm.2014.03.015
preparation, e.g. manual extraction of trachea and bronchi. The aim of our study is to evaluate some of the common emphysema quantification methods in a group of COPD patients in our community, to determine whether a simple semi-quantitative method aided by software enhanced visualization can be reliable enough for the complete evaluation of emphysema in COPD patients during the daily workload of a busy radiology department.

2. Patients and methods

Thirty patients (20 males and 10 females; mean age: 60.2 ± 4.4 years old) diagnosed clinically as having COPD were included in our study. The inclusion criteria were as follows: (1) ability to have a standard MDCT study of the chest. (2) CT findings of normal or emphysematous changes only were accepted (see #1 in exclusion criteria for explanation). (3) Complete pulmonary function tests.

Exclusion criteria were as follows: (1) Patients with chest CT findings of consolidation, collapse, malignancy, or pleural abnormalities that might affect the total lung volume. (2) Patients with incomplete records. (3) Patients with respiratory failure.

Patients had a standard non-enhanced MDCT study of the chest using a 16 slice GE BrightSpeed machine (GE Healthcare Medical Systems, Milwaukee, WI). Before the MDCT study, patients were instructed that they will be required to take and hold deep inspiration upon request during the study. They were trained on this maneuver 5 min before the scanning starts. The parameters of the CT scan include: 120 kV, auto tube mAs, slice thickness 5 mm, inter-slice gap 5 mm and resolution 512 × 512.

All patients had complete pulmonary function tests and according to the FEV\textsubscript{1} and FEV\textsubscript{1}/FVC, they were classified into the respective GOLD stage (2). None of our patients were stage IV.

<table>
<thead>
<tr>
<th>Stage</th>
<th>COPD</th>
<th>FEV\textsubscript{1}/FVC &lt; 0.7</th>
<th>FEV\textsubscript{1} ≥80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage I</td>
<td>Mild</td>
<td>FEV\textsubscript{1}/FVC &lt; 0.7</td>
<td>FEV\textsubscript{1} ≥80%</td>
</tr>
<tr>
<td>Stage II</td>
<td>Moderate</td>
<td>FEV\textsubscript{1}/FVC &lt; 0.7</td>
<td>FEV\textsubscript{1} 50–70%</td>
</tr>
<tr>
<td>Stage III</td>
<td>Severe</td>
<td>FEV\textsubscript{1}/FVC &lt; 0.7</td>
<td>FEV\textsubscript{1} 30–40%</td>
</tr>
<tr>
<td>Stage IV</td>
<td>Very severe</td>
<td>FEV\textsubscript{1}/FVC &lt; 0.7</td>
<td>FEV\textsubscript{1} &lt; 30%</td>
</tr>
</tbody>
</table>

2.1. Image analysis

Emphysema is identified as areas of hypovascular low attenuation. The images were evaluated visually and scored according to the modified Goddard scoring system which states that: (1) no signs of emphysema (score 0). (2) Emphysema in ≤5% (score 0.5). (3) Emphysema in <25% (score 1). (4) Emphysema in 25–50% (score 2). (5) Emphysema in 51–75% (score 3). (6) Emphysema in ≥75% (score 4) (3). This scoring system was first done on the grey scale images. Then this scoring was repeated again after applying a density mask to the image sequence. The density mask is a density threshold (−950 to −1024 HU) that highlights voxels within this density range (4). This level was chosen because it correlated best to the emphysematous changes in the lungs (5). The trachea, mainstem bronchi, bowel gas and the background of the image are included in this density range and were not excluded as this was a visual score. The local software of the MDCT workstation (Advantage Windows 4.4 software, GE Healthcare Medical Systems, Milwaukee, WI) was used for the density mask application. Three independent radiologists evaluated the images in the same manner and on separate occasions and their readings were recorded. Repeatability of scoring was done by one of the authors (L.A.M.) on 2 separate occasions, 1 week apart.

For the quantitative evaluation, again the local software of the MDCT workstation (Advantage Windows 4.4 software) was used for the segmentation and CT emphysema index calculation. The CT emphysema index is defined as the proportion of the lung affected by emphysema (1,4). Segmentation of the lung was done as a prior step to exclude soft tissues and fat from the field of analysis. A threshold of −200 to −1024 HU was applied to the entire image sequence and the rest of tissues were excluded. Trachea, mainstem bronchi, bowel gas and the background of the images were excluded manually (6). The image sequence was then revised for correct segmentation. After image manipulation, the segmented image sequence is saved and transferred to another computer. Image J software (National Institutes of Health, USA) was used for the calculation of the mean lung attenuation (MLA) and the emphysema index. The images were quantitatively evaluated considering a density level of −950 HU as the threshold level for emphysema (5).

The extent of airway disease was not assessed in this study.

2.2. Statistical analysis

Statistical analysis was run on SPSS v16 for Windows (SPSS, Cary, NC). The tests included the kappa test for inter-rater reliability of visual analysis. Repeatability was tested by calculating the intra-class correlation coefficient (ICC) and only after applying the density mask to the images. Spearman bivariate correlation test was then used to flag the correlations with the emphysema index. Analysis of Variance (ANOVA) was used to compare the means of MLA and emphysema index between the GOLD stages. The median visual score (of the 3 readers) was recorded before and after applying the density mask. Kruskal–Wallis test was used to compare the visual score between the GOLD stages. Statistical significance was used at \(p \leq 0.05\).

3. Results

Thirty patients with clinically diagnosed COPD were included in this study. The mean age and standard deviation was 60.2 ± 4.4 years old. The study included 20 males and 10 females. Four patients had GOLD stage I, 18 patients had GOLD stage II while 8 patients had GOLD stage III.

3.1. Inter-rater agreement and repeatability

The modified Goddard scale used for visual analysis of emphysema varied between the 3 readers as follows (Table 1).

The kappa test run between the 3 readers revealed slight agreement \((k = 0.122, p = 0.001)\).

After density mask application, the visual analysis scoring varied as follows (Table 2).
The kappa test run between the 3 readers revealed substantial agreement ($k = 0.75, p < 0.0005$).

A high degree of repeatability was found between the 1-week apart scores, with a single class ICC of 0.93 with a 95% confidence interval of 0.86–0.97. The median difference of emphysema score was $0 \pm 0.5$.

### 3.2. Bivariate correlation

Using the Spearman bivariate correlation test, the median visual score before density mask application, correlated positively to the emphysema index ($r = 0.62, p = 0.013$). After density mask application, the median visual score showed a stronger correlation to the emphysema index ($r = 0.81, p < 0.0005$; this was at the $p = 0.01$ level).

### 3.3. Quantitative measures and GOLD stage

The relation of the visual score, emphysema score, MLA and emphysema index to the GOLD stage is represented in the next table (Table 3).

### 4. Discussion

COPD is a very common disease in developing as well as developed countries due to the different smoking habits as well as air pollution and occupational exposure. However, it is a heterogeneous disease and spirometric results alone failed to explain the heterogeneity of the disease (7,8).

The marked heterogeneity of the disease led to a problem with clinical classification, where none of the available clinical staging systems alone allow enough prognostic information and adequate follow up of the patient care (9). This caused a growing interest in using CT to give a phenotypic classification for COPD patients, that may aid this clinical characterization of patients. Fujimoto et al. proposed classifying COPD patients according to their CT findings into Phenotype A (no or minimal emphysema with or without airway disease), Phenotype E (emphysema without airway disease) and Phenotype M (mixed airway and emphysema disease) (1). A similar approach was also proposed by Han et al., where they suggested that CT could be used to discriminate between patients with the same spirometric results (10).

With the continuous development in disease management, researchers have developed new quantification methods for the evaluation of the different chest MDCT findings in COPD patients. Airway changes are quantified by several parameters, e.g. wall thickness, total airway count and square root of wall area. Airspace (emphysematous) changes are also quantified by some parameters, e.g. emphysema index and semi-quantitative visual scoring system. The airspace and the airway measures, when co-existent, negatively correlate in COPD patients (11,12).

Although the airspace damage is irreversible, many researchers have found that the pre-operative quantitative CT assessment of emphysema and its distribution predicted a better post-operative outcome. For example, some authors suggested that upper lobe emphysema had a better outcome after lung volume reduction surgery (LVRS) than patients with predominantly lower lobe emphysema (13–15). Other authors suggested that pre-operative CT measures of emphysema predicted a better cardio-pulmonary exercise capacity (16).

A common problem that is encountered with all quantitative analysis is the time factor. All quantitative measures require some image modification, e.g. segmentation, thresholding and manual extraction of some components that would otherwise affect the calculation process. Also, the overlap between the disease categories or stages hinders the presence of a reliable cut-off value that differentiates between the disease stages. On the other hand, qualitative measures are highly variable between raters and vary with their experience levels.

In our study, we present a simple method that may help reduce the time needed for quantification of images and provide a more reliable and fast semi-quantitative method, hoping that it may allow the routine quantification of emphysema in routine chest CT studies of COPD patients.

We used the modified Goddard semi-quantitative score and combined it with a density mask that is available on most commercial workstations (3). This density mask technique is also available on many open source image processing software. This density mask step is a simple preliminary step that precedes the calculation of emphysema index on most software.

### Table 1  The variable scores given to the patients based on the modified Goddard score by the 3 independent readers (Before density mask application).

<table>
<thead>
<tr>
<th>Reader 1</th>
<th>Reader 2</th>
<th>Reader 3</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>6</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2  The variable scores given to the patients based on the modified Goddard score by the 3 independent readers (after density mask application).

<table>
<thead>
<tr>
<th>Reader 1</th>
<th>Reader 2</th>
<th>Reader 3</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>8</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.5</td>
<td>2</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We wanted to test the inter-rater variability before and after this density mask application. Inter-rater agreement increased from slight (\(k = 0.122\)) to substantial agreement (\(k = 0.75\)). Without the density mask, the visual score was variably over or under-estimated (Figs. 1–3). Few authors have resorted to this density mask method (4,17). In the study by Müller et al., this method showed good correlation to the pathological severity of emphysema (17). Müller et al. used the density mask for quantitative assessment of the extent of emphysema and to the authors' best knowledge, it was not reported whether combining this method with a visual score for emphysema would have a good inter-rater agreement than if it was used on grey-scale images. The modified Goddard visual score also proved good correlation to the CT quantitative measures of emphysema in a previous study (18). We had similar results, when we compared the median score of the 3 readers before and after the density mask application to the emphysema index. We had a stronger correlation between them after its application than before (\(r = 0.81, \ p < 0.0005\) vs \(r = 0.62, \ p = 0.013\), respectively). Testing the repeatability of the visual score, we also had a high degree of reliability (single measure ICC = 0.93). To the authors' best knowledge, repeatability was not tested previously for any emphysema visual scoring system.

We then evaluated the visual score and the quantitative parameters between the available GOLD stages of COPD patients, in our study. Although all measures of emphysema were increasingly getting severe from stage I to stage III, there was significant overlap between the different stages regarding the visual score, emphysema index and mean lung attenuation. This is probably due to the low number of cases. The positive correlation between the quantitative emphysema measures (emphysema index and MLA) and the clinical stage was proved in many studies before (19,20).

Many authors failed to find a direct relation between the emphysema index and the clinical stage of the patients and again, they attributed this to the heterogeneity of COPD and the associated airway disease (18,21–24). Positive relation was found when patients with Phenotype E were selectively studied, where this group of patients had lower diffusing capacity for carbon monoxide (\(DL_{CO}\)) and no responsiveness to bronchodilators (1,25). The distribution of emphysema also shows different relations to pulmonary function, where patients with predominant upper lobe emphysema had lower FEV\(_1\), FEV\(_1\)/FVC and FVC after follow up compared to patients with predominant lower lobe emphysema (26). Accordingly, the phenotype of COPD on CT greatly affects the pulmonary functions.

There are some limitations in our study. First, the low number of patients is definitely a drawback and it did not allow us to study the relation between the density mask aided score and the pulmonary function tests adequately. However, the relation of the visual score to the pulmonary functions was not part of the aim of our study and it is already well known that the relation would be poor, because we did not select a certain phenotype of COPD.

Also, the low number of patients was probably the reason for the absence of correlation between the MLA and the GOLD stage, a fact that was proven in many studies (19,20). However, another explanation could be the fact that the used clinical staging and the inspiratory–expiratory state of the patient during the MDCT examination differed between these studies and ours.

Second, the CT scan used in this study is a standard non contrast enhanced examination. It is well known that the pre-determined threshold levels vary with the CT acquisition parameters, e.g. slice thickness, intra-venous contrast as well as the reconstruction algorithm used (27–30). Although with

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GOLD I (n = 4)</th>
<th>GOLD II (n = 18)</th>
<th>GOLD III (n = 8)</th>
<th>Significance level (ANOVA and K–W tests)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLA</td>
<td>(-748 \pm 40.8)</td>
<td>(-761.6 \pm 41.2)</td>
<td>(-781.3 \pm 39.8)</td>
<td>0.61</td>
</tr>
<tr>
<td>Emphysema index</td>
<td>2.63 ± 2.36</td>
<td>6.48 ± 7.33</td>
<td>7.74 ± 3.53</td>
<td>0.65</td>
</tr>
<tr>
<td>Visual score before (range)</td>
<td>0–0.5</td>
<td>0–3</td>
<td>0–2</td>
<td>0.19</td>
</tr>
<tr>
<td>Visual score after (range)</td>
<td>0.5</td>
<td>0.5–3</td>
<td>1–3</td>
<td>0.04</td>
</tr>
</tbody>
</table>

![Fig. 1](A) A patient with COPD. (A) CT image of the chest reveals normal lung parenchyma. Before density mask application, the visual score given to this patient was 0. (B) After density mask application, the visual score was 0.5. Although the emphysematous region is small, it affected the score.
modern CT scanners, these effects are minimized and the acquisition parameters, used in this study, are nearly universal and used in everyday practice, none of our patients had intra-venous contrast. We did not study whether the score would differ if patients were given intra-venous contrast or not. Accordingly, the results of our study will only be applicable to non-enhanced CT examination.

Finally, we did not include patients with CT findings of consolidation, collapse, malignancy or pleural abnormalities. Although it is not uncommon to see these pathological entities in COPD patients, they may alter the lung volumetry and emphysema index. We believe that these changes would not alter the visual scoring of emphysema which could be another advantage for the visual scoring. Yet this was not tested in our study.

5. Conclusion

COPD is a heterogeneous disease with 3 main CT derived phenotypes. Within each phenotype, it is essential to have a reliable and valid method for assessment of radiological changes, which in association with other pulmonary function tests, can aid the follow up of the appropriate treatment given to patients. Emphysema is one of the COPD phenotypes, either alone (Phenotype E) or mixed with airway disease (Phenotype M). Emphysema index was proved in many studies to predict clinical outcome, mortality and post-operative outcome after LVRS (11,15,21). In patients with emphysema, although the emphysema index is an accurate and sensitive method for quantifying emphysema, yet as any quantitative measure, it requires time-lengthy image preparation (19).

We present a simple visual score for the quantitation of emphysema, that when combined with a simple density mask, the inter-rater agreement and repeatability of the score are markedly improved. We did not choose any particular phenotype of COPD as this was not the aim of our study. This scoring method appears to be fast and easy to perform during the daily reporting activity in a busy radiology department.

Funding

No funding issues to disclose. This work did not receive any funding.

Conflict of interest

The authors report no conflict of interest.

Fig. 2 A patient with COPD. (A) CT image of the chest reveals emphysematous changes in both lungs. Before density mask application, the visual score given to this patient was 2. (B) After density mask application, the visual score was 1. Some areas that were considered normal on the first image appeared emphysematous on the second image (white arrows) and vice versa with other regions (yellow arrows).

Fig. 3 A patient with COPD. (A) CT image of the chest reveals normal lung parenchyma. Before density mask application, the visual score given to this patient was 0. (B) After density mask application, the visual score was 0.5.
Agreement

All authors have materially participated in this work and agree to the submission.

References


