Decreasing Variability in CT Scan Reporting of Emphysema

FL Jacobson, MD, MPH; R Madan, MD; RR Gill, MBBS; O Buckley, MBBCh; H Hatabu, MD, PhD; AR Hunsaker, MD

Division of Thoracic Imaging, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

PURPOSE

This project provides documented practice quality improvement required by the American Board of Radiology for maintenance of certification. It provides standards, tools and feedback mechanisms to support life-long learning and improve patient care. Phenotypes of COPD classified purely based on severity of emphysema are not well defined and may be different from classical phenotypes that are currently being studied as part of the COPDGene™ research program. A quality assurance project comparing clinical and research interpretations for consistency has led to incorporating the COPD-Gene™ consensus standard cases for education and guidance during interpretation of CT scans. The initial objective was to reduce intra- and inter-observer variability. The resulting quality assurance and improvement project is designed for ongoing feedback and adaptation to revisions of the standard images and novel image displays.

BACKGROUND

The pulmonary division of the department of medicine at Brigham and Women's Hospital would like to use the COPD Gene™ CT protocol for clinical evaluations of COPD over time. This would require greater consistency of reporting of emphysema from very large datasets including ultrathin section reconstruction of inspiratory and reduced dose expiratory CT scans. This provided the motivation for developing greater agreement between subspecialty-trained radiologists, including those who did not participate as readers in the National Emphysema Treatment. The scoring of emphysema, primarily performed for research studies, had not been systematically taught to cardiothoracic imaging fellows. The establishment of reference standards first used at the COPDGene™ Workshop at the ACR Learning Center in Reston, VA provided the needed educational resources and comparison standards to perform this Quality Assurance/Quality Improvement project.

The National Emphysema Treatment Trial adapted visual scoring of emphysema on CT images to a zonal system applied to non-contiguous HRCT images and through a training set of images increased the consistency of reporting between radiologists. The film-based reference standards were no longer available when the COPDGene™ commenced. The reporting of emphysema by zone without such standardization resulted in greater variability in both the general reporting of emphysema and the specific scores.

The Global Initiative for Chronic Obstructive Lung Disease (GOLD) was established in 2001 to promote research in COPD. GOLD COPD classifications describe the severity of the obstruction or airflow limitation. Four categories of severity for COPD are based on the value of FEV1 that tends to decrease with disease progression.

STAGE	COPD SEVERITY	FEV1/FVC	FEV1
I	Mild	<0.70	≥ 80% normal
II	Moderate	<0.70	50-79% normal
III	Severe	<0.70	30-49% normal
IV	Very Severe	<0.70	<30% normal, or <50% normal with chronic respiratory failure present*

* Usually requiring long-term oxygen therapy

METHODS

COPDGene™ study patient CT scans received clinical reports in accordance with the radiology department policy, generally within 24 hours. Meanwhile, the interpreting thoracic radiologist also provided data entry for the COPDGene™ study that included emphysema zonal distribution and severity. The visual scoring system followed procedures previously used for the National Emphysema Treatment Trial with 4 categories, 1-25%/26-50%/51-75%/>75%, for each zone. Forty-five COPDGene™ subject CT scans were randomly selected to represent GOLD Stages 0-4 (10 in each Stage 0-3 and 5 Stage 4). Radiology reports were reviewed and compared with the emphysema assessment performed by the same radiologist. Standardized images, presented at the COPDGene™ Workshop, were subsequently placed on a shared network drive for access in the reading room and can be easily revised without creating confusion as these standards continue to evolve. Each radiologist reviewed the standards as an educational activity prior to scoring emphysema. The images can also be freely accessed while performing emphysema scoring. For each subject, two different series were selected for these readings, a clinical presentation of 1 x 10 mm axial HRCT images in B60f lung algorithm and the COPDGene™ research image series of 0.75 x 0.60 mm images in B41f algorithm. The presentations were randomized together for 6 radiologists to score emphysema on five of the 90 presentations each week on an ongoing basis. Intra- and inter-observer variability assessments will continue to be provided at intervals. Statistical analyses include Fisher's exact test. Over time, additional cases and different display presentations will be evaluated in the same manner without overwhelming the clinical service. The provision of feedback is central to the ongoing nature of this quality assurance project.

We established the BWH COPDGene™ case library by randomly selecting cases within each GOLD COPD stage classification enrolled in the COPDGene™ Trial at Brigham and Women's Hospital. The library is thus biased to earlier emphysema that is less consistently scored. The CT scan images, up to approximately 3000 per case, have been stored on the BWH APPI-server for more rapid access than possible when retrieving previously reported cases from the BWH Clinical PACS. The emphysema score reported to the COPDGene™ data center and the clinical report were retrieved and compared.

Practice-Based Learning and Improvement Project for Maintenance of Certification

Quality Assurance
- QA comparison
- Clinical reading of emphysema
- Emphysema score

Training Set
- COPDGene™ standard images
- Score emphysema research and clinical issues

Quality Improvement
- Feedback
- Decreased variability in emphysema score
- Adapt to changes in standards and images

25% Emphysema

50% Emphysema

75% Emphysema

Zone Reference

RESULTS

The variability of clinical reports has been quantified and used to determine the area of greatest variability. Clinical and research readings agree for 27 of 45 CT scans. 11 scans were scored as having no emphysema, although the absence of emphysema was offered as a pertinent negative less than 50% of the time. Paraseptal emphysema alone was included in score by one radiologist and excluded from score by another radiologist. Emphysema scores less than 6 were associated with equivocation in the clinical report revealing the underlying uncertainty regarding the range of normal. Clinical descriptors varied most for moderate emphysema in 13 scans. The same reader was found to assign moderate and severe to the same score in different patients. One scan that had no mention of emphysema in the clinical report received an emphysema score of 24, the highest score for any of these scans. The most complete clinical reports had the highest specific correlation with emphysema score at virtually 100%. Of note, the grouping of scores shows separation between mild, moderate and severe despite the inconsistencies of descriptive reporting.

Emphysema was not consistently included in the clinical report, even when significant emphysema score was greater than 10, although inconsistency was greater for mild emphysema. Of note, emphysema was present in GOLD Stage 0 cases.

Parameter	Clinical Report	Emphysema Score Range
No Emphysema	11	0-2
Emphysema	34	1-24
Mild	6	1-6
Moderate	13	8-15
Severe	6	18-24

FELLOWSHIP TRAINING

The image standards for 25%, 50%, and 75% emphysema from the COPDGene Imaging Workshop were reviewed with the fellow along with a coronal image of the chest landmarks used for volumetric division of lungs into thirds whereby the upper zone extends to the carina rather than the aortic arch. This change did not result in differences in zonal scoring within the BWH COPDGene™ case library. The likelihood of less clarity in lobar attribution, as in the case of upper lobe predominant emphysema, and the implications of that description were also reviewed with the fellow as part of the training. After reviewing the standards, groups of 5 randomly chosen scans were scored. By the second group, the fellow had a working understanding of emphysema scoring that could be independently applied with greater consistency than in the original data collected. The standard images are therefore recommended for use at the time of emphysema scoring.

DISCUSSION - Issues in Ongoing QA

The review of cases needs to be limited to not more than 5 within the same interval of time. It would be preferable to have 1 case added to a clinical work-list each day to allow all of the radiologists to score the same 5 randomly chosen cases each week. The recent insertion of mammograms into the clinical reading stream is being used as a model in this regard.

Repeated readings of cases by all of the radiologists in the division decreases the variability in reporting emphysema descriptively and increases the consistent use of scoring system. The ability to also disseminate changes, such as the change in landmark used for division between thirds of the lung, support this project as a framework for continuing professional education as required for maintenance of certification. This strategy can be employed to increase agreement between readers for any image evaluation task including tasks that will be created in the future.

CONCLUSION

Our data collection provides a pathway to decreasing intra- and inter-observer variability and a robust method for comparing the effect of image presentation in the subjective evaluation of emphysema. The systematic incorporation of reference standard images for comparison will be helpful for teaching of residents and fellows. This methodology will also help a busy clinical service select the most reliable image presentation for a particular clinical decision.

REFERENCES

1. Hersh CP, Washko G, Jacobson FL, Gill R, Estepar RSJ, Reilly JJ, Silverman EK. Interobserver variability in the determination of upper lobe-predominant emphysema. Chest 2007; 131(2):424-431.
2. Lynch DA, Crapo JD, Murphy JR, Jacobson FL, Newell JD, Cauczor H. Patterns of COPD on CT: Consensus cases from the COPDGene(TM) workshop presented on Bbehalf of the COPDGene (TM) workshop participants. Education Exhibit at 2010 RSNA, *LL-CHE-MO7B*.
3. Nishimura K, Murata K, Yamagishi M, Itoh H, Ikeda A, Tsukino M, Koyama H, Sakai N, Mishima M, Izumi T. Comparison of different computed tomography scanning methods for quantifying emphysema. J Thorac Imaging 1998;13:193–198.
4. Global Strategy for Diagnosis, Management and Prevention of COPD – Updated 2005. Available from http://www.goldcopd.org