

UC San Diego

UC San Diego Previously Published Works

Title

Reader Perceptions and Impact of AI on CT Assessment of Air Trapping

Permalink

<https://escholarship.org/uc/item/06k2861n>

Journal

Radiology Artificial Intelligence, 4(2)

ISSN

2638-6100

Authors

Retson, Tara A
Hasenstab, Kyle A
Kligerman, Seth J
et al.

Publication Date

2022-03-01

DOI

10.1148/ryai.2021210160

Peer reviewed

Reader Perceptions and Impact of AI on CT Assessment of Air Trapping

Tara A. Retson, MD, PhD* • Kyle A. Hasenstab, PhD* • Seth J. Kligerman, MD • Kathleen E. Jacobs, MD • Andrew C. Yen, MD • Sharon S. Brouha, MD • Lewis D. Hahn, MD • Albert Hsiao, MD, PhD

From the Department of Radiology, University of California, San Diego, 9452 Medical Center Dr, 4th Floor, La Jolla, CA 92037 (T.A.R., S.J.K., K.E.J., A.C.Y., S.S.B., L.D.H., A.H.); and Department of Mathematics and Statistics, San Diego State University, San Diego, Calif (K.A.H.). Received June 15, 2021; revision requested August 23; revision received September 22; accepted October 22. Address correspondence to T.A.R. (e-mail: tretson@health.ucsd.edu).

*T.A.R. and K.A.H. contributed equally to this work.

Supported by research grants from Microsoft AI for Health and Amazon Web Services. T.A.R. supported by the National Institutes of Health (T32 EB005970), the Radiological Society of North America (grant RR1879), and the Friedman Family Endowed Radiology Fellowship. A.H. supported by the National Science Foundation (grant no. 2026809).

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2021; 4(2):e210160 • <https://doi.org/10.1148/ryai.2021210160> • Content codes: **AI** **CH** **CT**

Quantitative imaging measurements can be facilitated by artificial intelligence (AI) algorithms, but how they might impact decision-making and be perceived by radiologists remains uncertain. After creation of a dedicated inspiratory-expiratory CT examination and concurrent deployment of a quantitative AI algorithm for assessing air trapping, five cardiothoracic radiologists retrospectively evaluated severity of air trapping on 17 examination studies. Air trapping severity of each lobe was evaluated in three stages: qualitatively (visually); semiquantitatively, allowing manual region-of-interest measurements; and quantitatively, using results from an AI algorithm. Readers were surveyed on each case for their perceptions of the AI algorithm. The algorithm improved interreader agreement (intraclass correlation coefficients: visual, 0.28; semiquantitative, 0.40; quantitative, 0.84; $P < .001$) and improved correlation with pulmonary function testing (forced expiratory volume in 1 second-to-forced vital capacity ratio) (visual $r = -0.26$, semiquantitative $r = -0.32$, quantitative $r = -0.44$). Readers perceived moderate agreement with the AI algorithm (Likert scale average, 3.7 of 5), a mild impact on their final assessment (average, 2.6), and a neutral perception of overall utility (average, 3.5). Though the AI algorithm objectively improved interreader consistency and correlation with pulmonary function testing, individual readers did not immediately perceive this benefit, revealing a potential barrier to clinical adoption.

© RSNA, 2021

Air trapping and bronchiolitis obliterans are important pathways for diffuse lung injury, caused by inflammatory obstruction of small airways. They are seen in a variety of respiratory conditions, including chronic obstructive pulmonary disease (COPD) (1), graft-versus-host disease, bronchiolitis obliterans syndrome (BOS), chronic rejection or allograft dysfunction in lung transplant patients (2,3), cystic fibrosis, and hypersensitivity pneumonitis, among several others (3–5). These share common manifestations on chest CT scans, observed as a mosaic pattern of lung attenuation and loss of change in parenchymal attenuation between inspiratory and expiratory phase images (6). Measures of air trapping can also be used to evaluate effects of treatment and long-term progression, particularly in COPD and lung transplantation (7–9). While radiologist interpretation of chest CT scans remains the clinical standard, air trapping detection is a visually difficult task (7). Evidence shows potential for quantitative CT methods to detect and grade characteristics such as air trapping on a granular level (1,10) and to prognosticate outcomes in disease processes like BOS (11–13).

Quantitative measurements show benefit but can be time and labor intensive (14). Machine learning algorithms leveraging convolutional neural networks have the potential to automate quantitative measurements (15,16) and may make them more feasible clinically. Despite their capabilities, it remains uncertain how to translate these

algorithms into clinical practice, how they might practically affect diagnostic decision-making, and how they might be perceived in the radiologist workflow.

To evaluate the impact and reader perceptions of a quantitative artificial intelligence (AI) algorithm on the assessment of air trapping severity, we deployed a previously developed algorithm to quantify air trapping into our clinical practice alongside a dedicated quantitative inspiratory-expiratory chest CT examination (17). Five subspecialty cardiothoracic radiologists retrospectively reviewed these examinations, assessed algorithm impact at multiple stages of interpretation, and were surveyed about their perceptions of the algorithm after each case.

Materials and Methods

In this retrospective study, 18 consecutive quantitative inspiratory-expiratory CT scans were obtained, excluding one nondiagnostic examination due to insufficient inspiration. Examinations were performed at our institution between December 2020 and March 2021 with non-contrast-enhanced images obtained at full inspiration and normal expiration. CT examinations were performed with one of three CT scanners, including a GE Revolution 256 and GE 64 (GE Healthcare) and Canon 320 (Canon Medical Systems). Thin-section 0.5-mm or 0.625-mm images were reconstructed with standard reconstruction kernels and were provided for reader review and automated analysis with the AI algorithm.

Abbreviations

AI = artificial intelligence, BOS = bronchiolitis obliterans syndrome, COPD = chronic obstructive pulmonary disease, FEV₁ = forced expiratory volume in 1 second, FEV_{1pp} = FEV₁ percent predicted, FVC = forced vital capacity, PFT = pulmonary function test, ROI = region of interest

Summary

Quantitative artificial intelligence (AI) measurement of air trapping on CT scans increased interreader consistency and improved correlation with pulmonary function testing; however, reader perception of utility did not immediately align with the objective benefits, highlighting a potential barrier to AI adoption.

Key Points

- Quantitative artificial intelligence (AI) measurements improved interreader consistency in assessment of air trapping on dedicated inspiratory-expiratory chest CT scans (intraclass correlation coefficient, 0.28 to 0.84).
- Quantitative AI measurements improved the correlation between assessment of air trapping and forced expiratory volume in 1 second-to-forced vital capacity ratio on dedicated inspiratory-expiratory chest CT scans ($r = -0.26$ to -0.44).
- Reader's subjective impression of algorithm utility did not always align with the objective benefits, indicating a disconnect between interpretative value and reader perception.

Keywords

Technology Assessment, Quantification

Automated Analysis of Air Trapping

Images were analyzed with an in-house-developed algorithm, described in detail in Hasenstab et al, to perform lobar-level quantification of emphysema and air trapping (17). This algorithm consists of a series of convolutional neural networks that perform lobar-level lung segmentation and deformable registration between inspiratory and expiratory phases. This algorithm generates a table of measurements and a color overlay image series to highlight areas of emphysema and air trapping.

Image Review by Cardiothoracic Subspecialty Radiologists

Five fellowship-trained cardiothoracic radiologists (S.J.K., K.E.J., A.C.Y., S.S.B., L.D.H., average of 7 years of postfellowship experience) assessed air trapping severity for each lung lobe on inspiratory and expiratory images. Severity categories were based on prior observations from the COPDGene study cohort (17) as follows: (a) normal, less than 15% air trapping; (b) mild, 15%–33%; (c) moderate, 33%–50%; (d) severe, 50%–66%; and (e) very severe, greater than 66%. Readers reviewed images in three phases (Fig 1). First, readers rated the severity of air trapping visually, without use of region-of-interest (ROI) tools or results of the AI algorithm. Second, readers rated air trapping semiquantitatively, allowing use of an ROI tool displaying lung attenuation in Hounsfield units. We informed the readers that the algorithm utilized an attenuation difference threshold of 100 HU to establish the presence or absence of air trapping, in addition to providing the scientific publication describing the algorithm (17). Third, readers rated air trapping quantitatively, allowing readers to form a

final assessment either incorporating or discarding the algorithm's results. Image review resulted in 1275 score observations across the 17 patients \times five lung lobes \times five readers \times three methods (visual, semiquantitative, and quantitative). After each case, readers subjectively rated the following: their level of agreement with the algorithm, on a scale of 1 (completely disagree) to 5 (completely agree); the algorithm impact on their final assessment, on a scale of 1 (no change) to 5 (completely different); and how useful they found the algorithm, on a scale of 1 (not helpful) to 5 (extremely helpful).

Interreader agreement between severity categories was assessed with intraclass correlation coefficient (18), separately for each air trapping evaluation strategy (visual, semiquantitative, quantitative), using a cross-classified random-effects model with a random intercept corresponding to lung lobe nested within patient and a random intercept corresponding to readers. Ninety-five percent CIs were calculated using clustered bootstrapping with resampling performed at the patient level. Comparison of scoring differences between each of the three air trapping evaluation strategies was performed using a cross-classified mixed-effects model with a three-level categorical variable (visual, semiquantitative, or quantitative) as a fixed effect, lung lobe nested within patient as a random effect, and reader as a separate random intercept. Scoring differences and 95% CIs were determined analytically using the cross-classified mixed-effects model. Pulmonary function tests (PFTs) were available from the clinical record for 16 patients. To determine correlations between air trapping rating and forced expiratory volume in 1 second (FEV₁), FEV₁ percent predicted (FEV_{1pp}), and FEV₁-to-forced vital capacity (FVC) ratio for each evaluation strategy, scores were first averaged within patient and within reader, producing a single score per patient per reader. Correlations were then calculated between the patients' averaged scores and PFT values within reader, producing five correlations per patient. Resulting correlations were then averaged across the five readers for the final correlation measure. Correlation 95% CIs and differences between strategies were determined using a clustered bootstrapping procedure with resampling at the patient level. Correlation between PFTs and the unadjusted quantitative AI results were calculated using a Pearson correlation. All statistical analysis was performed in R (R Foundation for Statistical Computing) using a type I error rate of .05 for statistical significance.

Results

Average patient age was 57 years (range, 24–86 years; eight women, nine men). For demographics and study indications, please see Table 1.

On average, the severity rating between visual, semiquantitative, and quantitative measures increased with use of ROI and AI-generated measurements. Comparing visual and semiquantitative assessment, there was a small but significant difference, with an increase of 0.21 severity grades ($P < .001$; 95% CI: 0.15, 0.28). When scores were averaged among all readers, four patients had a change greater than 0.5 grades; if rounded to nearest severity grade, this would have resulted in a higher severity rating. Comparing visual and quantitative assessment, there was an increase of 0.83 grades ($P < .001$; 95% CI: 0.71, 0.98), with

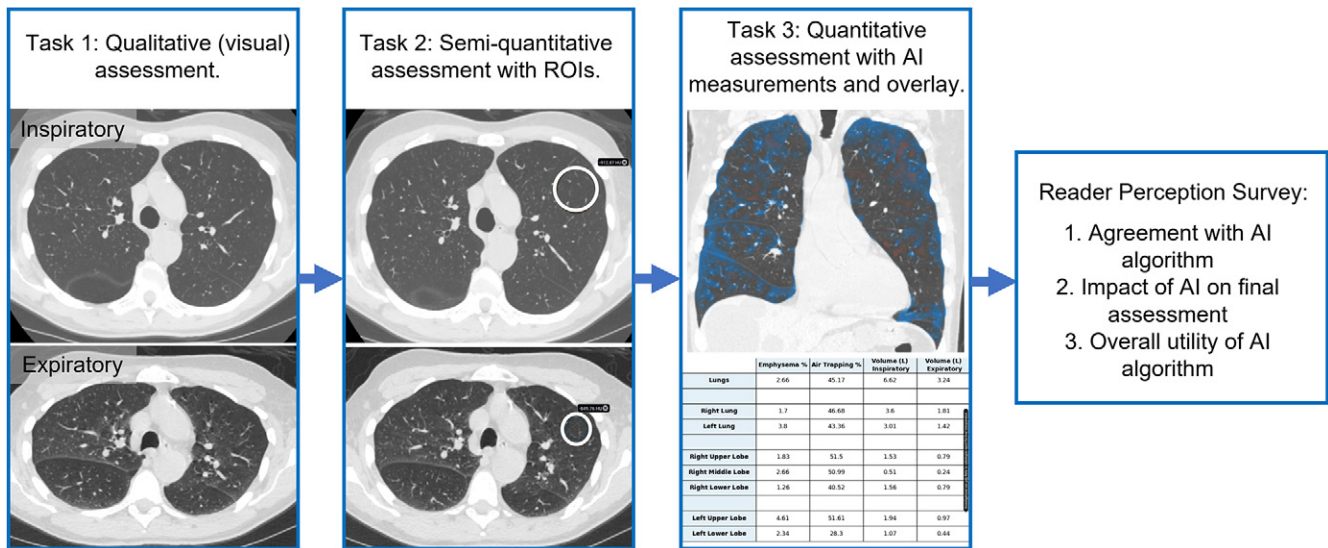


Figure 1: Design of the reader study. Cardiothoracic radiologists performed lobar-level assessment of air trapping severity in three stages. **1.** Air trapping severity was rated visually with inspiratory and expiratory images. **2.** Air trapping was assessed semiquantitatively after placement of any desired regions of interest (ROIs) to measure lung attenuation. **3.** Air trapping was rated after providing artificial intelligence (AI)-generated quantitative measurements and color overlays showing areas of air trapping (blue) and emphysema (red). Finally, readers were surveyed for their perceptions of the AI algorithm.

Table 1: Patient Demographics and Study Indications from the Clinical Record for Quantitative Inspiratory-Expiratory Lung CT

Patient Demographic and Indication	Value
Male-to-female ratio	9:8
Average age (y)	57 (range, 24–86)
Study indication	
Organ transplant	6
Marrow or stem cell transplant	4
Other bronchiolitis obliterans	2
Tracheobronchomalacia	2
Restrictive pathologic condition or PFT	2
Airways disease	1
Smoking status	
Never smoker	10
Former smoker	7
Average pack-years	17 (range, 2.5–58)

Note.—Except where otherwise noted, values are numbers of patients. No patients were current smokers. PFT = pulmonary function test.

11 patients having their rating changed to a higher severity, and between semiquantitative and quantitative there was an increase of 0.62 grades ($P < .001$; 95% CI: 0.52, 0.72), with 11 patients having their rating changed to a higher severity. A case where assessment of severity increased by nearly one grade between semiquantitative and quantitative analysis is highlighted in Figure 2. The patient later developed pneumomediastinum, a rare complication of BOS.

Intraclass correlation coefficient reader agreement increased with semiquantitative analysis and access to AI-generated quantitative measurements (Table 2). Visual agreement was low at 0.28 (95% CI: 0.12, 0.43), increased for semiquantitative at

0.40 (95% CI: 0.28, 0.51), and good for quantitative at 0.84 (95% CI: 0.78, 0.88).

Reader assessments were compared against measurements from PFTs, including FEV_1 , FEV_{1pp} , and FEV_1/FVC . There was modest correlation between assessments of air trapping to FEV_1 or FEV_{1pp} (Table 2). FEV_1/FVC correlation increased from $r = -0.26$ for visual assessment to $r = -0.32$ for semiquantitative and to $r = -0.44$ for quantitative. Ranges of Pearson correlations for PFTs across readers decreased considerably between all visual and quantitative assessments. However, differences in mean correlations were not significantly different from each other ($P = .25-.76$), likely related to sample size. All unadjusted AI quantitative measures fell within the ranges of quantitative reader scores for FEV_1 , FEV_{1pp} , and FEV_1/FVC . For FEV_1 , the raw AI correlation was slightly higher than the reader mean at $r = -0.42$, for FEV_{1pp} it was slightly lower at $r = -0.17$, and for FEV_1/FVC it was slightly higher at $r = -0.47$ (Table 2).

Reader perception of algorithm impact on assessment and overall utility was variable between readers, but generally favorable. On average, there was moderate perceived agreement with the algorithm, averaging 3.7 of 5 (range, 3.1–4.1). On average, perceived impact on final assessment was rated as little change, with average of 2.6 of 5 (range, 2.1–3.4), and perceived utility was rated slightly above neutral at 3.5 of 5 (range, 2.3–4.9).

Discussion

Here we observed the impact of an AI algorithm on subspecialty cardiothoracic radiologist interpretation of air trapping on inspiratory-expiratory lung CT scans. Quantitative AI measurements improved interreader consistency beyond manual ROI

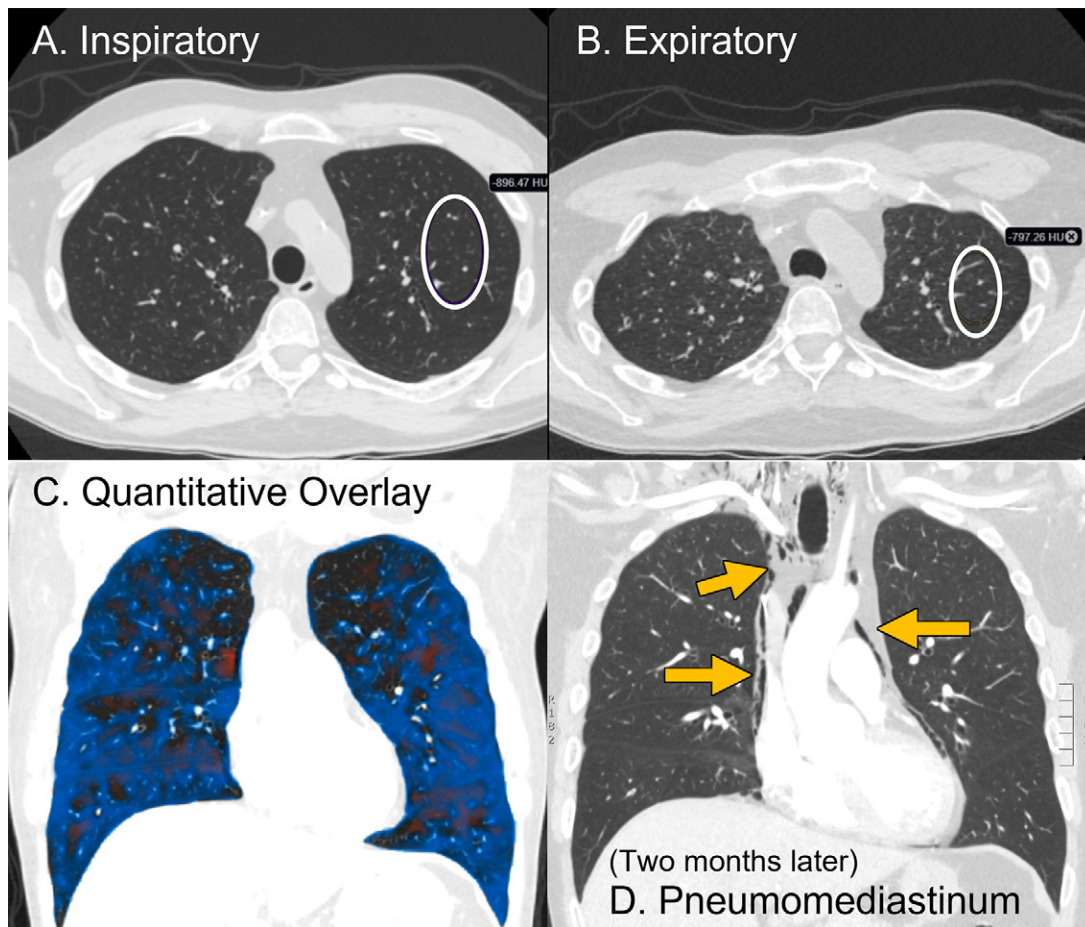


Figure 2: Example case in a 45-year-old man with history of stem cell transplant and graft-versus-host disease with bronchiolitis obliterans syndrome (BOS) causing diffuse air trapping. At (A) inspiration and (B) expiration, lung attenuation in the left upper lobe in the regions of interest (ROIs) were 896 HU and 797 HU, respectively. (C) On the artificial intelligence–generated quantitative overlay, there is extensive air trapping throughout the lungs (shown in blue, with areas of emphysema shown in red). For this case, readers’ assessment of air trapping increased by nearly one grade between placement of ROIs and provision of quantitative maps. (D) Two months after the initial CT, the patient went on to develop spontaneous pneumomediastinum (arrows), a rare complication of severe BOS.

measurements, and the algorithm improved correlation with objective measurements of pulmonary function. Acknowledging the reader sample size is small, qualitative perception of algorithm utility in this population did not immediately align with these objective benefits, indicating a disconnect between clinical value and reader perception.

In line with the benefits seen here, quantitative measurements can improve reliability of imaging interpretation but remain underutilized in many clinical scenarios. For example, while it is common practice to assess size and growth of pulmonary nodules with a single long-axis measurement, newer guidelines suggest nodule volume is more reliable (19). Quantitative volumetry improves nodule classification (20), reduces interreader variability (21), and allows calculation of volume doubling time, considered a better metric of nodule growth (2). Similarly, while pitting AI against radiologists has drawn considerable attention, multiple recent studies have shown that when AI is paired with a radiologist, the effect may be synergistic. In addition to several examples from the mammography literature (22,23), Liu et al found that a lung nodule detection algorithm performed similarly to

radiologists, but when paired with radiologists, their performance improved (24).

Previous work has shown correlations between air trapping and PFTs, with FEV_1 or FEV_{pp} in the range of $r = -0.45$ (25). However, in studies similar to ours with ranges of disease states and emphysema, the relationship with PFTs is less clear (26,27). FEV_1 was found to account for less than half of air trapping progression on CT scans in patients with COPD (28), and correlations are lower between AI measures and PFTs in cystic fibrosis (27). While PFTs allow comparison against an objective measurement, their variability in both acquisition and association with imaging findings highlights a need for future studies to determine their relationship to imaging across disease states and longitudinally.

There were several limitations to this study. First, radiologist perception was assessed early in the clinical deployment of the AI algorithm, a few months after creation of the dedicated quantitative lung CT examination. With greater experience, reader perception of the algorithm may change, or the algorithm may gradually influence reader interpretation. This type of automation bias is seen in a variety of fields with automated decision

Table 2: Interreader Reliability and Correlation to Pulmonary Function Testing for Each Method of CT Air Trapping Assessment

Parameter	Qualitative (Visual)	Semiquantitative (ROI)	Quantitative (AI-aided)	AI (Unadjusted)
Interreader reliability	0.28 ($P < .001$); 95% CI: 0.12, 0.43	0.40 ($P < .001$); 95% CI: 0.28, 0.51	0.84 ($P < .001$); 95% CI: 0.78, 0.88	...
Correlation with FEV ₁	Mean: -0.10 (-0.41 to 0.04)	Mean: -0.07 (-0.40 to 0.10)	Mean: -0.18 (-0.25 to -0.14)	-0.24
Correlation with FEV ₁ percent predicted	Mean: -0.13 (-0.34 to 0.09)	Mean: -0.11 (-0.19 to -0.03)	Mean: -0.20 (-0.24 to -0.17)	-0.17
Correlation with FEV ₁ /FVC	Mean: -0.26 (-0.66 to -0.10)	Mean: -0.32 (-0.55 to -0.06)	Mean: -0.44 (-0.50 to -.39)	-0.47

Note.—Except where otherwise noted, means with ranges are presented. Among five cardiothoracic radiologists, interreader reliability of assessments of air trapping improved with incorporation of semiquantitative and quantitative measurements, as did correlation with spirometric measurements of FEV₁/FVC. There was mild correlation with FEV₁ or FEV₁ percent predicted. AI = artificial intelligence, FEV₁ = forced expiratory volume in 1 second, FVC = forced vital capacity, ROI = region of interest.

support, ranging from aviation to health care, and has been shown to have a greater impact on less experienced physicians (29,30). Future work can assess longer-term impact on reader agreement with AI or patient outcomes. Further, different algorithms may not observe as dramatic an effect on reader agreement as seen in air trapping, a relatively difficult visual task compared with many other aspects of diagnostic imaging.

The benefit of AI-generated quantitative measurements on clinical interpretation of air trapping is highlighted here, as is a potential barrier to adoption of AI into clinical practice—reader perception. It may be difficult for an individual to perceive benefit, which may lag behind the larger impact. When seeking to implement AI algorithms in the clinical workflow, benchmarks of impact on interpretation, such as interreader agreement, or correlation with other objective external metrics relevant to the disease process, such as PFT testing, may be required.

Author contributions: Guarantors of integrity of entire study, T.A.R., K.A.H., A.H.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, T.A.R., S.J.K., A.H.; clinical studies, K.E.J.; experimental studies, T.A.R., K.A.H., S.J.K., S.S.B., L.D.H., A.H.; statistical analysis, T.A.R., A.H.; and manuscript editing, all authors

Disclosures of conflicts of interest: T.A.R. RSNA Machine Learning Committee member, unrelated to this work. K.A.H. No relevant relationships. S.J.K. Deputy editor of *Radiology: Cardiothoracic Imaging*. K.E.J. No relevant relationships. A.C.Y. No relevant relationships. S.S.B. No relevant relationships. L.D.H. No relevant relationships. A.H. Grants from GE Healthcare and Bayer; cofounder and shareholder in Arterys.

References

1. Lowe KE, Regan EA, Anzueto A, et al. COPDGene® 2019: Redefining the Diagnosis of Chronic Obstructive Pulmonary Disease. *Chronic Obstr Pulm Dis* (Miami) 2019;6(5):384–399.
2. Devaraj A, van Ginneken B, Nair A, Baldwin D. Use of Volumetry for Lung Nodule Management: Theory and Practice. *Radiology* 2017;284(3):630–644.
3. Miller WT Jr, Chatzkel J, Hewitt MG. Expiratory air trapping on thoracic computed tomography. A diagnostic subclassification. *Ann Am Thorac Soc* 2014;11(6):874–881.
4. Criado E, Sánchez M, Ramírez J, et al. Pulmonary sarcoidosis: typical and atypical manifestations at high-resolution CT with pathologic correlation. *RadioGraphics* 2010;30(6):1567–1586.
5. Hall GL, Logie KM, Parsons F, et al. Air trapping on chest CT is associated with worse ventilation distribution in infants with cystic fibrosis diagnosed following newborn screening. *PLoS One* 2011;6(8):e23932.
6. Mohamed Hoessein FA, de Jong PA. Air trapping on computed tomography: regional versus diffuse. *Eur Respir J* 2017;49(1):1601791.
7. Bin Saeedan M, Mukhopadhyay S, Lane CR, Renapurkar RD. Imaging indications and findings in evaluation of lung transplant graft dysfunction and rejection. *Insights Imaging* 2020;11(1):2.
8. de Jong PA, Dodd JD, Coxson HO, et al. Bronchiolitis obliterans following lung transplantation: early detection using computed tomographic scanning. *Thorax* 2006;61(9):799–804.
9. Pompe E, Strand M, Rikxoort EM, et al. Five-year Progression of Emphysema and Air Trapping at CT in Smokers with and Those without Chronic Obstructive Pulmonary Disease: Results from the COPDGene Study. *Radiology* 2020;295(1):218–226.
10. Chen A, Karwoski RA, Gierada DS, Bartholmai BJ, Koo CW. Quantitative CT Analysis of Diffuse Lung Disease. *RadioGraphics* 2020;40(1):28–43.
11. Verleden SE, Vos R, Vandermeulen E, et al. Parametric Response Mapping of Bronchiolitis Obliterans Syndrome Progression After Lung Transplantation. *Am J Transplant* 2016;16(11):3262–3269.
12. Van Herck A, Sacreas A, Heigl T, et al. Chest CT Has Prognostic Value at BOS Diagnosis after Lung Transplantation. *J Heart Lung Transplant* 2019;38(4 Suppl):S16–S17.
13. Gazourian L, Ash S, Meserve EEK, et al. Quantitative computed tomography assessment of bronchiolitis obliterans syndrome after lung transplantation. *Clin Transplant* 2017;31(5):e12943.
14. Suinesiaputra A, Cowan BR, Finn JP, et al. Left Ventricular Segmentation Challenge from Cardiac MRI: A Collation Study. Berlin, Germany: Springer, 2012; 88–97.
15. Schoppe O, Pan C, Coronel J, et al. Deep learning-enabled multi-organ segmentation in whole-body mouse scans. *Nat Commun* 2020;11(1):5626.
16. Zhang K, Liu X, Shen J, et al. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. *Cell* 2020;181(6):1423–1433.e11. [Published correction appears in *Cell* 2020;182(5):1360.]
17. Hasenstab KA, Yuan N, Retson T, et al. Automated CT Staging of Chronic Obstructive Pulmonary Disease Severity for Predicting Disease Progression and Mortality with a Deep Learning Convolutional Neural Network. *Radiol Cardiothorac Imaging* 2021;3(2):e200477.
18. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol* 2012;8(1):28–34.
19. Callister MEJ, Baldwin DR, Akram AR, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax* 2015;70(Suppl 2):ii1–ii54.
20. Mehta HJ, Ravenel JG, Shaftman SR, et al. The utility of nodule volume in the context of malignancy prediction for small pulmonary nodules. *Chest* 2014;145(3):464–472.
21. Jeon KN, Goo JM, Lee CH, et al. Computer-aided nodule detection and volumetry to reduce variability between radiologists in the interpretation of lung nodules at low-dose screening computed tomography. *Invest Radiol* 2012;47(8):457–461.

22. Schaffter T, Buist DSM, Lee CI, et al. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. *JAMA Netw Open* 2020;3(3):e200265.
23. Salim M, Wählin E, Dembrower K, et al. External Evaluation of 3 Commercial Artificial Intelligence Algorithms for Independent Assessment of Screening Mammograms. *JAMA Oncol* 2020;6(10):1581–1588.
24. Liu K, Li Q, Ma J, et al. Evaluating a Fully Automated Pulmonary Nodule Detection Approach and Its Impact on Radiologist Performance. *Radiol Artif Intell* 2019;1(3):e180084.
25. Arakawa H, Webb WR. Air trapping on expiratory high-resolution CT scans in the absence of inspiratory scan abnormalities: correlation with pulmonary function tests and differential diagnosis. *AJR Am J Roentgenol* 1998;170(5):1349–1353.
26. Stern EJ, Webb WR, Gamsu G. Dynamic quantitative computed tomography. A predictor of pulmonary function in obstructive lung diseases. *Invest Radiol* 1994;29(5):564–569.
27. Ram S, Hoff BA, Bell AJ, et al. Improved detection of air trapping on expiratory computed tomography using deep learning. *PLoS One* 2021;16(3):e0248902.
28. Pompe E, van Rikxoort EM, Schmidt M, et al. Parametric response mapping adds value to current computed tomography biomarkers in diagnosing chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2015;191(9):1084–1086.
29. Bond RR, Novotny T, Andrsova I, et al. Automation bias in medicine: The influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms. *J Electrocardiol* 2018;51(6S):S6–S11.
30. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc* 2012;19(1):121–127.