# UC San Diego
## UC San Diego Previously Published Works

**Title**
Automated CT Staging of Chronic Obstructive Pulmonary Disease Severity for Predicting Disease Progression and Mortality with a Deep Learning Convolutional Neural Network

**Permalink**

**Journal**

**ISSN**

**Authors**
Hasenstab, Kyle A
Yuan, Nancy
Retson, Tara
et al.

**Publication Date**

**DOI**

Peer reviewed

# Automated CT Staging of Chronic Obstructive Pulmonary Disease Severity for Predicting Disease Progression and Mortality with a Deep Learning Convolutional Neural Network

Kyle A. Hasenstab, PhD • Nancy Yuan, MS • Tara Retson, MD, PhD • Douglas J. Conrad, MD •
Seth Kligerman, MD • David A. Lynch, MB • Albert Hsiao, MD, PhD • for the COPDGene Investigators*

From the Department of Radiology (K.A.H., N.Y., T.R., S.K., A.H.) and Department of Medicine (D.J.C.), University of California San Diego, 9452 Medical Center Dr, La Jolla, CA 92037; Department of Mathematics and Statistics, San Diego State University, San Diego, Calif (K.A.H.); and Department of Radiology, National Jewish Health, Denver, Colo (D.A.L.). Received August 13, 2020; revision requested November 6; revision received January 29, 2021; accepted February 5. **Address correspondence to** K.A.H. (e-mail: *kahasenstab@sdsu.edu).*

*Members of the COPDGene study group can be found in the Appendix (supplement).

**Purpose:** To develop a deep learning–based algorithm to stage the severity of chronic obstructive pulmonary disease (COPD) through quantification of emphysema and air trapping on CT images and to assess the ability of the proposed stages to prognosticate 5-year progression and mortality.

**Materials and Methods:** In this retrospective study, an algorithm using co-registration and lung segmentation was developed in-house to automate quantification of emphysema and air trapping from inspiratory and expiratory CT images. The algorithm was then tested in a separate group of 8951 patients from the COPD Genetic Epidemiology study (date range, 2007–2017). With measurements of emphysema and air trapping, bivariable thresholds were determined to define CT stages of severity (mild, moderate, severe, and very severe) and were evaluated for their ability to prognosticate disease progression and mortality using logistic regression and Cox regression.

**Results:** On the basis of CT stages, the odds of disease progression were greatest among patients with very severe disease (odds ratio [OR], 2.67; 95% CI: 2.02, 3.53; *P* < .001) and were elevated in patients with moderate disease (OR, 1.50; 95% CI: 1.22, 1.84; *P* = .001). The hazard ratio of mortality for very severe disease at CT was 2.23 times the normal ratio (95% CI: 1.93, 2.58; *P* < .001). When combined with Global Initiative for Chronic Obstructive Lung Disease (GOLD) staging, patients with GOLD stage 2 disease had the greatest odds of disease progression when the CT stage was severe (OR, 4.48; 95% CI: 3.18, 6.31; *P* < .001) or very severe (OR, 4.72; 95% CI: 3.13, 7.13; *P* < .001).

**Conclusion:** Automated CT algorithms can facilitate staging of COPD severity, have diagnostic performance comparable with that of spirometric GOLD staging, and provide further prognostic value when used in conjunction with GOLD staging.

*Supplemental material is available for this article.*

©RSNA, 2021

*An earlier incorrect version of this article appeared online. This article was corrected on December 22, 2021.*

Chronic obstructive pulmonary disease (COPD) is histologically defined by chronic airway inflammation, destruction of downstream alveoli and vasculature, and hyperinflation as air becomes trapped behind the obstructed airways. COPD affects more than 16 million Americans and is the fourth leading cause of death in the United States behind heart disease, cancer, and accidental death (1). Although COPD can result from various toxic inhalations or asthma, it is most commonly secondary to cigarette smoking (2,3). Diagnosis often relies on history of tobacco use or second-hand exposure, symptoms, and pulmonary function testing. However, it is becoming apparent that quantitative CT measurements have potential

diagnostic and prognostic value (4,5), as they correlate with spirometry-based pulmonary function testing findings (6). As such, CT measurements have been included in the recently updated diagnostic criteria for COPD from the investigators of the National Institutes of Health–supported COPD Genetic Epidemiology (COPDGene) study (7).

Despite the emerging use of quantitative CT, stratification of patients suspected of having COPD is currently based on spirometric pulmonary function testing findings according to Global Initiative for Chronic Obstructive Lung Disease (GOLD) stages (8). Analogous staging criteria with CT have not yet been defined. As diagnostic criteria begin to incorporate CT, it remains

## Abbreviations

AUC = area under the receiver operating characteristic curve, CNN = convolutional neural network, COPD = chronic obstructive pulmonary disease, COPDGene = COPD Genetic Epidemiology, $FEV_1$ = forced expiratory volume in 1 second, $FEV_1pp$ = percentage predicted $FEV_1$, FVC = forced vital capacity, GOLD = Global Initiative for Chronic Obstructive Lung Disease, HR = hazard ratio, OR = odds ratio, %AT = percentage air trapping, %EM = percentage emphysema, %TLI = percentage total lung involvement, PRISm = preserved ratio impaired spirometry

## Summary

CT-based severity stratification of patients with chronic obstructive pulmonary disease can prognosticate disease progression and mortality and may allow for improved diagnosis and management.

## Key Points

- The combination of deformable registration and deep learning lung segmentation can be used to reliably automate quantification of air trapping and emphysema from inspiratory and expiratory CT; measurements of low-attenuating areas agree with third-party image analyses (intraclass correlation, >0.99; 95% CI: >0.99, 1.00).
- Quantitative measurements of air trapping and emphysema can be used to define CT stages of chronic obstructive pulmonary disease (COPD) severity, which parallel the spirometrically defined Global Initiative for Chronic Obstructive Lung Disease (GOLD) stages; the combination of air trapping and emphysema predicts the GOLD stage with an area under the receiver operating characteristic curve of 0.86–0.96.
- Proposed CT-based stages for COPD severity prognosticate spirometric disease progression and mortality when implemented independently or in conjunction with the spirometric GOLD criteria; patients with severe disease at CT have greater odds of disease progression (odds ratio, 1.50–2.67; $P < .001$) and a greater hazard of mortality (hazard ratio, 2.23; $P < .001$).

unclear how quantitative measurements might be used to establish disease severity in the clinic, and current determinations rely heavily on qualitative visual assessment (9,10). Air trapping in particular is an image feature associated with small airway disease and is a clinically important component of COPD (11,12). However, it is often overlooked, as it typically cannot be directly visualized like emphysema, bronchial wall thickening, or endobronchial mucus plugging from a single inspiratory phase. Meanwhile, deformable registration algorithms have been used to both quantify and visualize the distribution of air trapping (13–16), and recent innovations in deep convolutional neural networks (CNNs) have shown promise for automating and supplementing a variety of tasks in medical imaging (17–20), including automated lung and lung-lobe segmentation (21–24). In this study, we recognize that quantitative CT measurements, although explored in the research domain, have not yet permeated the clinical arena.

We therefore build on this prior work to assess the feasibility of defining CT-based stages of COPD severity on the basis of a combination of air trapping and emphysema data obtained from inspiratory- and expiratory-phase CT. With the hypothesis that these two CT metrics alone might provide clinical prognostic value, we leveraged deep learning and deformable registration techniques to automate quantification

of emphysema and air trapping across patients in the COPDGene study. We used these image features to propose CT-based stages for disease severity. We then assessed the ability of our CT-based criteria to prognosticate disease progression and mortality.

## Materials and Methods

### Patient Data Sets

This retrospective Health Insurance Portability and Accountability Act–compliant study was approved by the institutional review boards of the participating institutions, which waived the requirement for written informed consent. There were two distinct sets of patients included in this study: one for development of a lung-segmentation CNN and a second using this CNN and deformable registration to define CT-based stages of COPD severity.

*Data set for CNN model development.—* We retrospectively collected a convenience sample of 1037 volumetric CT series from 888 current or former smokers undergoing lung cancer screening using low-radiation-exposure protocols to develop our own deep learning–based lung segmentation algorithm. Studies included 621 noncontrast CT examinations performed as part of the National Lung Screening Trial (25) and 416 noncontrast CT examinations performed at our institution between August 2008 and October 2018. Imaging data were anonymized prior to CNN training; therefore, demographic information is not available. Imaging protocols and parameters are provided in Table E1 (supplement).

*Data set for CT-based staging of COPD severity.—* To develop a CT-based strategy for staging COPD severity, we used baseline noncontrast CT and spirometric data from all 9652 patients with CT examinations from the COPDGene study (26). The COPDGene project has resulted in almost 400 publications from other research groups (27). To our knowledge, the analysis presented in this article, which defines CT-based criteria for staging COPD, has not been previously performed on the COPDGene data set. Exclusion criteria were missing inspiratory or expiratory CT studies (*n* = 536) and missing all-cause mortality status (*n* = 165), resulting in 8951 patients who were included in this study. Patients from this data set were used for testing of the CNN and derivation of stages of disease severity.

Spirometric measurements comprised the forced expiratory volume in 1 second ($FEV_1$), percentage predicted $FEV_1$ ($FEV_1pp$), and forced vital capacity (FVC) after administration of 180 μg of albuterol at baseline and at 5-year follow-up (7). Baseline CT images and spirometric measurements were collected between 2007 and 2011, and 5-year follow-up measurements were collected between 2013 and 2017. Data also included 10-year all-cause mortality, determined through longitudinal monitoring of patients every 6 months by the COPDGene investigators since 2009. Patients with an $FEV_1$/FVC ratio of less than 0.7 were classified using the GOLD staging

### Table 1: Summary of Patient Characteristics of Sample Taken from COPDGene Study Cohort

| Characteristic | Value |
| --- | --- |
| Demographics (*n* = 8951) | |
| Men (%) | 53.35 (4775/8951) |
| Race (% NHW) | 68.04 (6090/8951) |
| Age (y) | 60 ± 9 |
| Body mass index (kg/m²) | 28.8 ± 6.2 |
| No. of pack-years | 44.2 ± 24.9 |
| Current smokers (%) | 52.54 (4703/8951) |
| Spirometry (*n* = 8895)* | |
| FEV$_1$pp | 77.3 ± 27.0 |
| FVCpp | 89.2 ± 18.1 |
| FEV$_1$/FVC | 0.65 ± 0.17 |
| GOLD 0 | 3855 |
| GOLD 1 | 695 |
| GOLD 2 | 1723 |
| GOLD 3 | 1022 |
| GOLD 4 | 529 |
| PRISm | 1071 |
| Imaging (*n* = 8951) | |
| %EM | 6.2 ± 9.8 |
| %AT | 41.5 ± 20.4 |
| %TLI | 47.9 ± 23.6 |
| Normal | 3278 |
| Mild | 1405 |
| Moderate | 1456 |
| Severe | 1175 |
| Very severe | 1637 |

Note.—Categorical values are shown as the percentage, with the number of patients in parentheses, and continuous variables are shown as the mean ± standard deviation. GOLD stage 0 is defined by patients who have an FEV$_1$/FVC ratio greater than or equal to 0.7 and an FEV$_1$pp of greater than or equal to 80. %AT = percentage air trapping, %EM = percentage emphysema, COPDGene = Chronic Obstructive Pulmonary Disease Genetic Epidemiology, GOLD = Global Initiative for Chronic Obstructive Lung Disease, FEV$_1$ = forced expiratory volume in 1 second, FEV$_1$pp = percentage predicted FEV$_1$, FVC = forced vital capacity, FVCpp = percentage predicted FVC, NHW = non-Hispanic White, PRISm = preserved ratio impaired spirometry, %TLI = percentage total lung involvement.
* A total of 56 patients had missing data for spirometric analysis.

system for COPD severity; patients with an FEV$_1$/FVC ratio greater than 0.7 and an FEV$_1$pp greater than 80% were used as the normal spirometric reference group (GOLD stage 0) (8). Other patients had airflow obstruction (FEV$_1$pp <80%) and reduced FVC at spirometry, leading to a preserved FEV$_1$/FVC ratio (>0.7). These patients, comprising 12% of the COPD-Gene cohort, were classified as having preserved ratio impaired spirometry (PRISm) results (28) and were not included in the analysis of COPD status for this study. Demographic data are provided in Table 1.

## CT Image Acquisition

CT scans for the patients within the COPDGene cohort were acquired using multidetector General Electric, Philips, or Siemens CT scanners, each with at least 16 detector channels. Images comprised inspiratory (200 mAs) and expiratory (50 mAs) acquisitions of the entire thorax without contrast material. Images were reconstructed using a standard soft-tissue kernel and used submillimeter section thicknesses (range, 0.625–0.9 mm) and intervals (range, 0.45–0.625 mm) with smooth and edge-enhancing algorithms. Further details about the COPDGene study design and imaging protocols are described by Regan et al (26).

## Quantitative Analysis of Emphysema and Air Trapping

The analysis pipeline, shown in Figure 1, combines our deep learning lung-segmentation algorithm (K.A.H.) with deformable registration to automate quantification of emphysema and air trapping. Details on the development and validation of the lung segmentation algorithm are described in Appendix E1 (supplement). Expiratory series were deformably registered to inspiratory series using a symmetric diffeomorphic registration algorithm (29,30). Lungs masks were then generated from inspiratory images using our lung segmentation CNN. Pixelwise subtractions of co-registered inspiratory and expiratory attenuation were then used to compute attenuation difference maps.

We desired to distinctly separate regions of the lungs affected by emphysema from those affected by air trapping using inspiratory images and attenuation difference maps. The percentage emphysema (%EM) was computed as the percentage of the lung on the inspiratory images with attenuation less than or equal to −950 HU (31,32); the percentage air trapping (%AT) was computed as the percentage of voxels with attenuation differences of less than or equal to 100 HU and inspiratory attenuation of greater than –950 HU. The attenuation difference map cutoff of 100 HU was determined empirically (Appendix E2 [supplement]). We then defined the percentage total lung involvement (%TLI) as the total percentage of the lung affected by either emphysema or air trapping (%TLI = %EM + %AT). Note that our approach for quantification of air trapping is similar to that of parametric response maps (13,14). An auxiliary analysis showing the advantages of our approach in the context of GOLD stage prediction is included in Appendix E3 (supplement).

## Selection of Thresholds to Define CT-based Stages of COPD

GOLD stages were converted into four categories: GOLD stages 1–4 versus not, GOLD stages 2–4 versus not, GOLD stages 3–4 versus not, and GOLD stage 4 versus not. We then performed a search of thresholds on the %EM and %TLI image features that jointly maximized the Youden index for predicting these GOLD categories. Thresholds were then used to guide the proposal of five CT-based stages of COPD: normal, mild, moderate, severe, and very severe. Note that we intentionally defined CT-based stages in this way to parallel the widely used functional GOLD stage classes. Details on the threshold search are described in Appendix E4 (supplement).
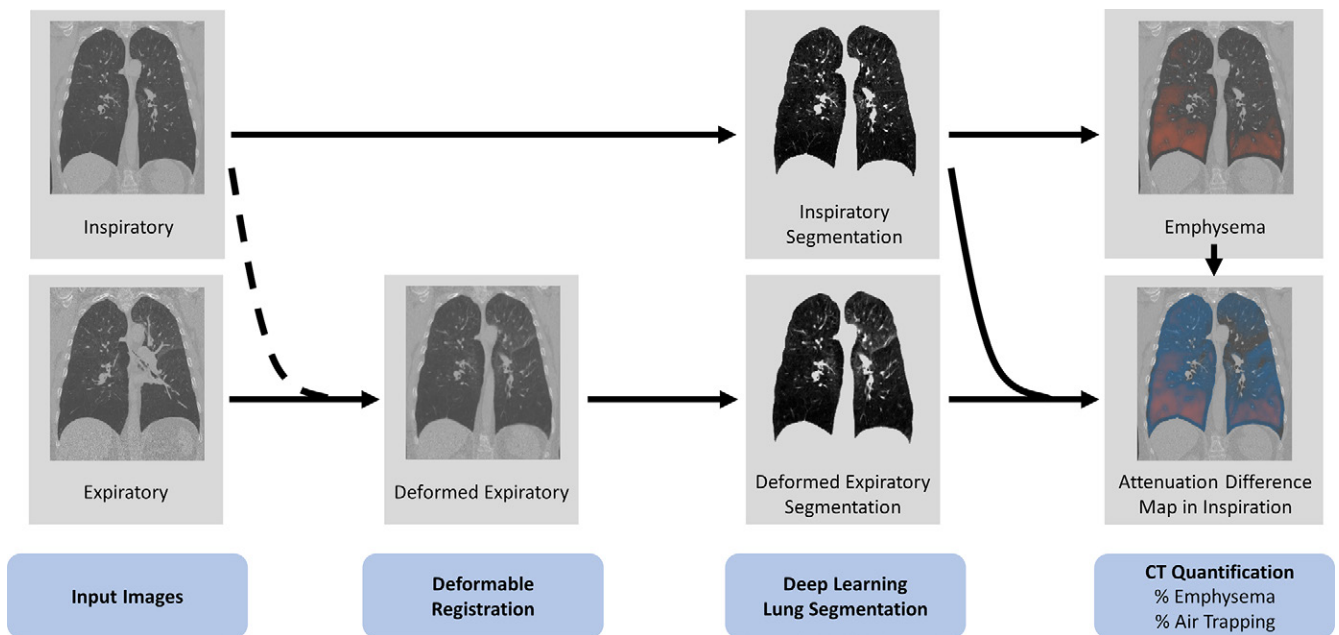
**Figure 1:** Pipeline for computing pathophysiologic image features using deep learning. Expiratory series were first deformably registered to inspiratory series for each patient. A three-dimensional deep learning lung segmentation algorithm was then used to segment the lungs from inspiratory images. Regions of emphysema were computed from inspiratory images (highlighted in red). The attenuation difference map was computed by subtracting inspiratory and deformed expiratory images (highlighted in blue). Maps were overlaid to spatially visualize total lung involvement. Dashed line indicates implicit use of the inspiratory series to produce the deformed expiratory series.

### Prognostic Evaluation of Proposed CT-based Stages of COPD

We assessed the ability of the GOLD stage and proposed CT-based stages to prognosticate two clinical outcomes: abnormal spirometric progression of $FEV_1$ ($FEV_1$ loss >350 mL between the baseline measurement and 5-year follow-up) and 10-year all-cause mortality. Additional details on the definition of $FEV_1$ progression and all-cause mortality can be found in Lowe et al (7). Detailed summaries of mortality by year, disease, and medications are included in Tables E3–E5 (supplement).

### Statistical Analysis

Deformable registrations and lung segmentations were performed in Python, version 3.6.2, using Diffusion Imaging in Python, version 0.14.0 (29), and Keras, version 2.2.0 (33), libraries. Statistical analysis was performed using R software (version 3.4.0; R Foundation for Statistical Computing). Patient demographics, image features, and spirometric metrics were summarized using descriptive statistics. We used uni- and multivariable logistic regression to predict the GOLD stage using the %EM and %AT as inputs and evaluated performance using the area under the receiver operating characteristic curve (AUC). Multivariable logistic regression was used to estimate the odds of $FEV_1$ loss across GOLD- and CT-based staging, adjusting for patient age at study entry, sex, race, current smoking status, number of pack-years, and baseline postbronchodilator $FEV_1$. We used Cox regression with right censoring to estimate hazard ratios (HRs) for all-cause mortality across GOLD- and CT-based staging, adjusting for sex, race, number of pack-years, and current smoking status; age was used as the underlying time scale. Sensitivity

and specificity and CIs for odds ratios (ORs) and HRs, were analytically calculated, as appropriate. Regression $P$ values were adjusted using Bonferroni correction; statistical significance was determined using a 5% type 1 error threshold.

## Results

### Patient Overview of Spirometry, GOLD Stage, and CT Characteristics

Spirometric measurements and GOLD stage are summarized in Table 1. A total of 56 patients had missing spirometric measurements and were omitted from analyses involving GOLD stages. GOLD stage 0 contained the largest number of patients (43%, 3855 of 8895); however, the mean $FEV_1pp$ (77.3% ± 27.0) and $FEV_1/FVC$ (0.65 ± 0.17) values were less than the thresholds ($FEV_1pp$ = 80% and $FEV_1/FVC$ = 0.70) separating GOLD stage 0 from other GOLD stages, reflecting the distributional left skewness of the spirometric measurements. GOLD stage 4 contained the fewest patients (6%, 529 of 8895). Quantitative measures across the entire group of 8951 patients showed a mean %EM of 6.2 ± 9.8, a mean %AT of 41.5 ± 20.4, and a mean %TLI of 47.9 ± 23.6 (Table 1).

### CNN Model Performance

After training and validation of the lung segmentation CNN with the separate data set of 888 patients (Appendix E1 [supplement]), the model was then tested on the 8951 patients from the COPDGene data set. Scatterplots visualizing the agreement between CNN metrics and semiautomated software metrics from three separate image analysis platforms for inspiratory and expiratory series are shown in Figures E1 and E2 (supplement), respectively. Intraclass correlation coefficients
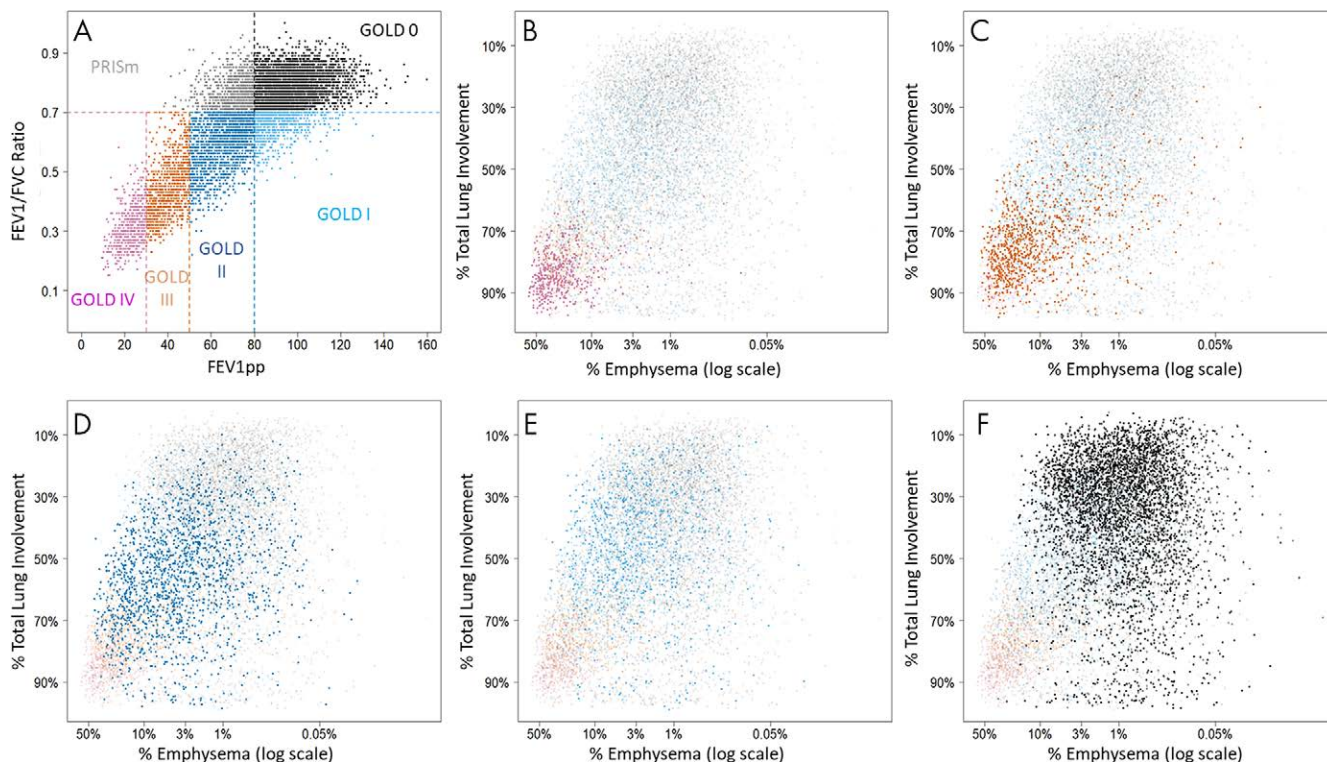
**Figure 2:** Relationship between Global Initiative for Chronic Obstructive Lung Disease (GOLD) stages derived using spirometric testing and CT image features. *A,* GOLD stages defined by the forced expiratory volume in 1 second ($FEV_1$)/forced vital capacity (FVC) ratio and percentage predicted $FEV_1$ ($FEV_1$pp) spirometric measurements. *B–F,* Two deep learning–derived pathophysiologic features, the percentage total lung involvement (y axes) and the percentage emphysema (x axes) are color coded by GOLD stage: *B,* GOLD stage 4; *C,* GOLD stage 3; *D,* GOLD stage 2; *E,* GOLD stage 1; and *F,* GOLD stage 0. Note that the axes of CT image features are log transformed and reversed for the purposes of visualization and analogy with spirometry. PRISm = preserved ratio impaired spirometry.

measuring the agreement between CNN metrics and semiautomated software metrics across the entire sample are shown in Table E2 (supplement). There was very strong agreement (intraclass correlation coefficient >0.99) across all validation metrics. The mean attenuation intraclass correlation coefficients were slightly smaller in magnitude than other metrics; however, the intraclass correlation coefficient among the three image analysis software platforms (0.98 [95% CI: 0.99, 0.99]) showed a similar decrease in agreement (Appendix E1 [supplement]). Total lung capacity, functional residual capacity, and mean attenuation slightly underestimated the semiautomated software metrics ($P < .001$), but the bias was not clinically meaningful. Although we did not have access to the original masks used to create the semiautomated software metrics, we hypothesize that this small bias is due to the inclusion of high-attenuating vasculature within the original ground truth masks used by COPDGene.
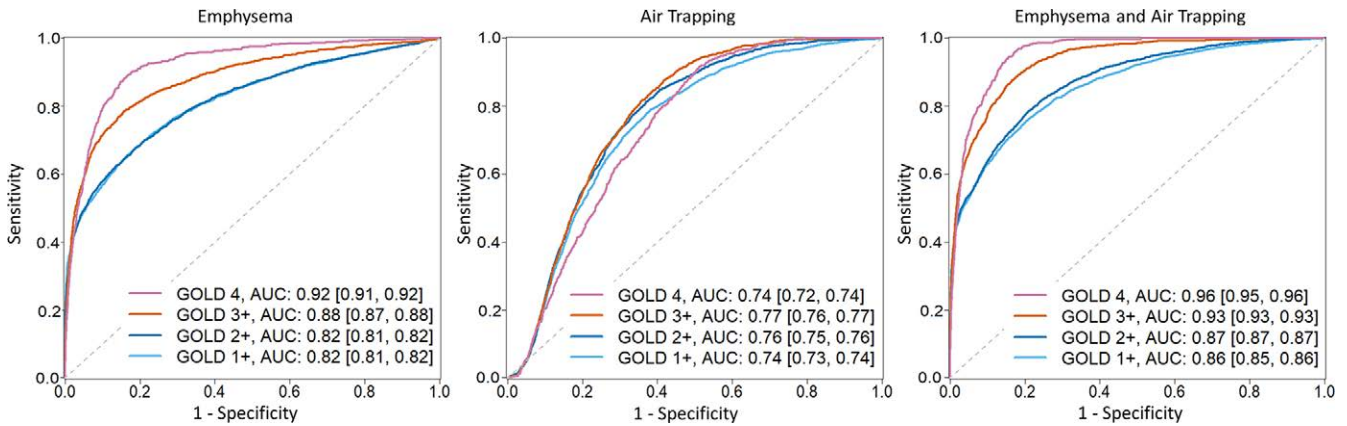
### GOLD Stage Prediction Using Image Features

Scatterplots of %TLI versus %EM color coded by GOLD stage (Fig 2, *A–F*) show the relationship between spirometry and image features. Plots indicate that the severity of the %EM and %TLI tends to increase with an increasing GOLD stage, as expected. Receiver operating characteristic curves for GOLD stage prediction for the %EM and %AT are shown in Figure 3. The %EM had strong predictive performance, with AUCs exceeding 0.82 for GOLD stages 1–3 and exceeding 0.92 for

GOLD stage 4. Conversely, the %AT weakly predicted the GOLD stage (AUCs < 0.80). However, AUCs using both the %EM and the %AT produced larger AUCs across GOLD categories *(P < .001)*.

### Image Thresholds and the Proposed CT-based Staging System

Thresholds for the %EM and %TLI and corresponding sensitivities and specificities for each GOLD category are shown in Table 2. In the bivariable model, thresholds were observed across a narrow range for the %EM (range, 1.19%–3.64%) and were observed across a wider range for the %TLI (range, 37.72%–67.88%). The %EM and %TLI alone were unable to distinguish between GOLD stages 1–4 and GOLD stages 2–4, as evidenced by their approximately equal image thresholds. Bivariable thresholds achieved greater Youden values than those achieved when using the %EM or %TLI individually. Sensitivities increased with the GOLD stage severity in both uni- and bivariable models.

Bivariable thresholds from Table 2 were used to guide the proposed definition of the CT-based COPD stages described in Table 3. CT-based stages applied to the study cohort are shown in Figure 4, *A.* The GOLD 1–4 and GOLD 2–4 categories are captured by the same moderate disease stage at CT because of their similar thresholds. Severe and very severe disease at CT capture the GOLD stage 3–4 and GOLD stage 4 categories, respectively. Additional image thresholds of 1%

**Figure 3:** Performance of logistic regression using CT image features to predict the Global Initiative for Chronic Obstructive Lung Disease (GOLD) stage. Note that confidence bands are excluded for visualization purposes. The receiver operating characteristic curves and areas under the receiver operating characteristic curve (AUCs) with 95% CIs in brackets for *A*, emphysema alone; *B*, air trapping alone; and *C*, emphysema and air trapping together for predicting the GOLD stage.

### Table 2: Uni- and Bivariable Empirical Thresholds for Pathophysiologic CT Features to Predict GOLD Stage

| Model by GOLD Stage | Sensitivity (%) | Specificity (%) | %EM | %TLI |
|---|---|---|---|---|
| Univariable, %EM* | | | | |
| 1+ | 65 (2593/3969) | 84 (4114/4926) | 3.12 | NA |
| 2+ | 67 (2198/3274) | 82 (4593/5621) | 3.53 | NA |
| 3+ | 78 (1217/1551) | 84 (6188/7344) | 5.60 | NA |
| 4 | 87 (462/529) | 85 (7128/8366) | 9.78 | NA |
| Univariable, %TLI* | | | | |
| 1+ | 77 (3058/3969) | 73 (3616/4926) | NA | 44.68 |
| 2+ | 80 (2635/3274) | 74 (4182/5621) | NA | 48.08 |
| 3+ | 91 (1407/1551) | 80 (5851/7344) | NA | 59.58 |
| 4 | 96 (507/529) | 82 (6827/8366) | NA | 69.00 |
| Bivariable, %EM and %TLI† | | | | |
| 1+ | 73 (2888/3969) | 86 (4234/4926) | 1.19 | 37.72 |
| 2+ | 78 (2553/3274) | 84 (4709/5621) | 1.19 | 40.78 |
| 3+ | 87 (1356/1551) | 87 (6410/7344) | 1.19 | 59.58 |
| 4 | 93 (492/529) | 88 (7376/8366) | 3.64 | 67.88 |
| Proposed thresholds‡ | | | | |
| 1+ | 85 (3364/3969) | 56 (2738/4926) | 1 | 10 |
| 2+ | 83 (2724/3274) | 74 (4164/5621) | 1 | 33 |
| 3+ | 91 (1404/1551) | 81 (5973/7344) | 1.5 | 50 |
| 4 | 94 (495/529) | 87 (7240/8366) | 3 | 66 |

Note.—Empirical thresholds were determined by maximizing the Youden index with deep learning–derived features. Youden thresholds were used as a guide in the determination of CT-based chronic obstructive pulmonary disease stages. GOLD = Global Initiative for Chronic Obstructive Lung Disease, NA = not applicable, %EM = percentage emphysema, %TLI = percentage total lung involvement.
* Univariable analyses use either emphysema only or total lung involvement only.
† Bivariable analysis used both emphysema and total lung involvement.
‡ Proposed CT-based thresholds are as follows: GOLD stage 1+ = mild, GOLD stage 2+ = moderate, GOLD stage 3+ = severe, and GOLD stage 4 = very severe.

for the %EM and 10% for the %TLI were separately selected to represent populations with minimal lung involvement and were used to define the normal imaging results reference category. In results similar to those for GOLD stage 0, we found

that a larger proportion of patients (37%, 3278 of 8951) were classified as having normal CT findings. The remaining CT-based stages contained an approximately equal proportion of patients (mild, 16% [1405 of 8951]; moderate, 16% [1456

**Table 3: Proposed CT-based COPD Stages**

| Stage | Percentage Emphysema | Operator | Percentage Total Lung Involvement |
|---|---|---|---|
| Normal | <1 | Or | <10 |
| Mild | >1 | And | 10–33 |
| Moderate | 1–1.5 | Or | 33–50 |
| Severe | 1.5–3 | Or | 50–66 |
| Very severe | >3 | And | >66 |

Note.—Thresholds were used as a guide to propose five CT-based stages of COPD that paralleled the Global Initiative for Chronic Obstructive Lung Disease staging system. COPD = chronic obstructive pulmonary disease.
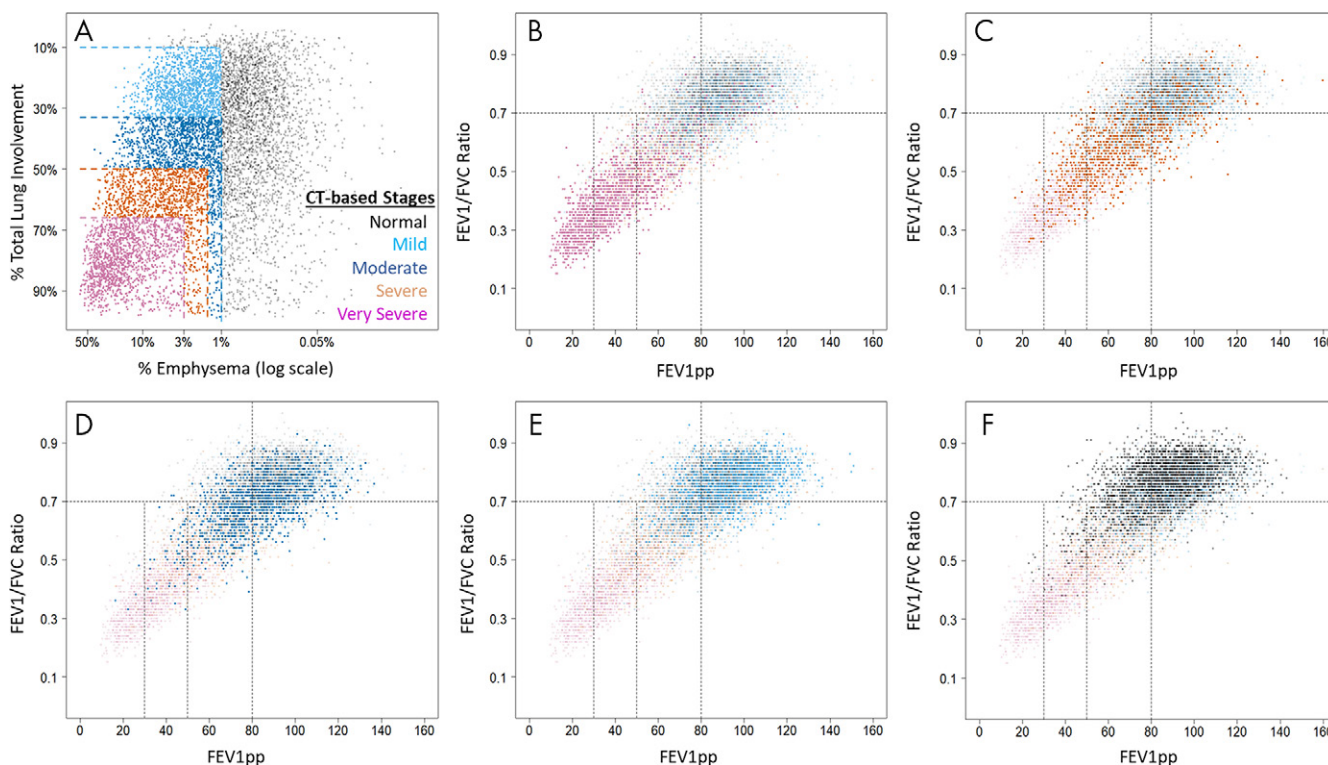


**Figure 4:** Relationship between proposed CT-based stages and spirometric pulmonary function testing results. *A*, Proposed imaging-based chronic obstructive pulmonary disease stages are defined on the basis of the percentage emphysema and the percentage total lung involvement. *B–F*, Global Initiative for Chronic Obstructive Lung Disease (GOLD) stages are color coded by the proposed CT-based stages: *B*, very severe; *C*, severe; *D*, moderate; *E*, mild; and *F*, normal. As expected, each of the proposed CT-based stages predicts spirometric severity. Patients with very severe disease at CT span GOLD stage 2–4 disease. Patients with severe disease at CT span GOLD stage 1–4 disease and include some patients with normal spirometric results. Many patients with moderate and mild disease at CT (dark blue) have normal spirometric results. $FEV_1$ = forced expiratory volume in 1 second, $FEV_1pp$ = percentage predicted $FEV_1$, FVC = forced vital capacity.

of 8951]; severe, 13% [1175 of 8951]; and very severe, 18% [1637 of 8951]).

Scatterplots of the $FEV_1$/FVC ratio versus the $FEV_1pp$ color coded by the proposed CT-based COPD stages (Fig 4, *A–F*) show the relationship between proposed CT-based stages and spirometric findings. As expected, each of the proposed CT-based stages predicts spirometric severity. However, patients with severe or very severe results at imaging largely span GOLD stages 2–4 or GOLD stages 1–3, respectively, with several of these patients showing normal spirometric findings. Normal-to-moderate imaging findings are largely contained within the GOLD 0 category.

### Prognostication of Clinical Outcomes

ORs and HRs for $FEV_1$ progression and all-cause mortality are summarized in Table 4 and Figure 5. Regression results for the PRISm group are reported separately in Table E6 (supplement). In the imaging-only reference category, the odds of disease progression for patients with moderate COPD (OR, 1.50 [95% CI: 1.22, 1.84]), severe COPD (OR, 2.88 [95% CI: 2.29, 3.63]), and very severe COPD (OR, 2.67 [95% CI: 2.01, 3.53]) were greater than those for the normal imaging results reference category ($P < .001$ for all). Patients with severe or very severe disease at CT had greater odds of progression than patients with mild or moderate disease, suggesting that the
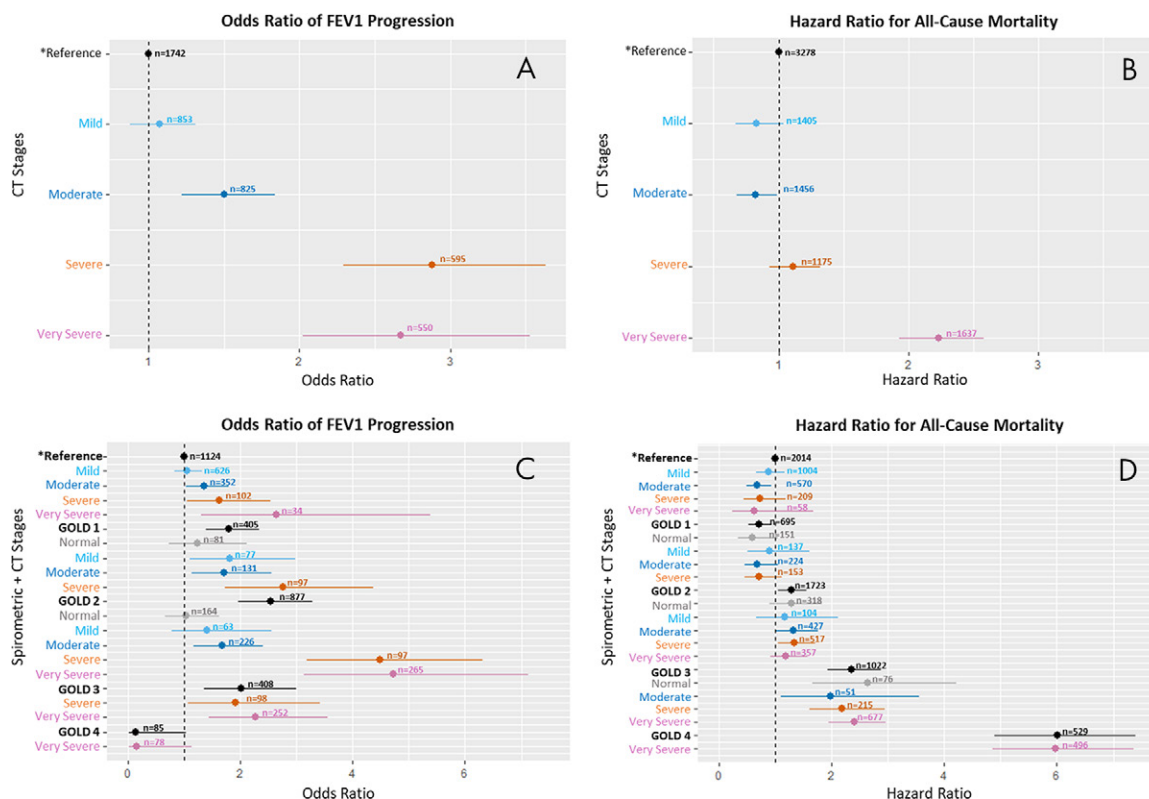
**Table 4: Predictive Ability of Imaging and Spirometry for COPD Progression and All-Cause Mortality**

| Characteristic | FEV$_1$ Progression | | | All-Cause Mortality | | |
| --- | --- | --- | --- | --- | --- | --- |
| | No. of Patients | Odds Ratio | Adjusted P Value | No. of Patients | Hazard Ratio | Adjusted P Value |
| Imaging only | | | | | | |
| None | 1742 | 1 | | 3278 | 1 | |
| Mild | 853 | 1.07 (0.88, 1.31) | >.99 | 1405 | 0.83 (0.67, 1.04) | .81 |
| Moderate | 825 | 1.50 (1.22, 1.84)* | .001 | 1456 | 0.82 (0.68, 0.99) | .29 |
| Severe | 595 | 2.88 (2.29, 3.63)* | <.001 | 1175 | 1.11 (0.93, 1.32) | >.99 |
| Very severe | 550 | 2.67 (2.02, 3.53)* | <.001 | 1637 | 2.23 (1.93, 2.58)* | <.001 |
| Spirometry only | | | | | | |
| GOLD stage 0 | 2238 | 1 | | 3855 | 1 | |
| GOLD stage 1 | 405 | 1.53 (1.20, 1.95)* | .006 | 695 | 0.83 (0.64, 1.06) | >.99 |
| GOLD stage 2 | 877 | 2.01 (1.55, 2.61)* | <.001 | 1723 | 1.51 (1.29, 1.78)* | <.001 |
| GOLD stage 3 | 408 | 1.47 (0.97, 2.24) | .69 | 1022 | 2.80 (2.38, 3.30)* | <.001 |
| GOLD stage 4 | 85 | 0.10 (0.01, 0.72) | .23 | 529 | 7.15 (6.03, 8.46)* | <.001 |
| Imaging and spirometry | | | | | | |
| GOLD stage 0 | | | | | | |
| None | 1124 | 1 | | 2014 | 1 | |
| Mild | 626 | 1.04 (0.82, 1.31) | .99 | 1004 | 0.87 (0.65, 1.17) | >.99 |
| Moderate | 352 | 1.36 (1.03, 1.80) | .79 | 570 | 0.68 (0.49, 0.93) | .46 |
| Severe | 102 | 1.63 (1.05, 2.54) | .79 | 209 | 0.72 (0.43, 1.19) | >.99 |
| Very severe | 34 | 2.64 (1.30, 5.38) | .19 | 58 | 0.62 (0.23, 1.67) | >.99 |
| GOLD stage 1 | | | | | | |
| None | 81 | 1.23 (0.72, 2.12) | >.99 | 151 | 0.59 (0.33, 1.04) | >.99 |
| Mild | 77 | 1.81 (1.10, 2.97) | .49 | 137 | 0.90 (0.51, 1.60) | >.99 |
| Moderate | 131 | 1.70 (1.13, 2.55) | .27 | 224 | 0.68 (0.45, 1.04) | >.99 |
| Severe | 97 | 2.75 (1.73, 4.36)* | <.001 | 153 | 0.71 (0.45, 1.12) | >.99 |
| Very severe | 19 | 1.37 (0.50, 3.76) | >.99 | 30 | 0.65 (0.21, 2.05) | >.99 |
| GOLD stage 2 | | | | | | |
| None | 164 | 1.03 (0.65, 1.62) | >.99 | 318 | 1.28 (0.90, 1.82) | >.99 |
| Mild | 63 | 1.40 (0.77, 2.56) | >.99 | 104 | 1.17 (0.65, 2.11) | >.99 |
| Moderate | 226 | 1.68 (1.17, 2.41) | .13 | 427 | 1.32 (0.99, 1.76) | >.99 |
| Severe | 265 | 4.48 (3.18, 6.31)* | <.001 | 517 | 1.33 (1.04, 1.71) | .63 |
| Very severe | 159 | 4.72 (3.13, 7.13)* | <.001 | 357 | 1.19 (0.91, 1.57) | >.99 |
| GOLD stage 3 | | | | | | |
| None | 32 | 0.49 (0.11, 2.13) | >.99 | 76 | 2.64 (1.65, 4.22)* | .001 |
| Mild | 2 | 0.00 (0.00, ∞) | >.99 | 3 | 5.29 (0.74, 37.93) | .99 |
| Moderate | 24 | 3.05 (1.17, 7.94) | .62 | 51 | 1.97 (1.09, 3.56) | .64 |
| Severe | 98 | 1.91 (1.07, 3.41) | .733 | 215 | 2.18 (1.61, 2.94)* | <.001 |
| Very severe | 252 | 2.26 (1.44, 3.55)* | .001 | 677 | 2.40 (1.94, 2.96)* | <.001 |
| GOLD stage 4 | | | | | | |
| None | 0 | | | 8 | 10.47 (4.61, 23.74) | <.001 |
| Mild | 0 | | | 0 | | |
| Moderate | 1 | 0 (0, ∞) | >.99 | 4 | 14.01 (3.46, 56.67) | .006 |
| Severe | 6 | 0 (0, ∞) | >.99 | 21 | 5.09 (2.67, 9.71) | <.001 |
| Very severe | 78 | 0.15 (0.02, 1.14) | >.99 | 496 | 5.98 (4.86, 7.36)* | <.001 |

Note.—Hazard ratios and odds ratios are shown with 95% CIs. COPD = chronic obstructive pulmonary disease, FEV$_1$ = forced expiratory volume in 1 second, GOLD = Global Initiative for Chronic Obstructive Lung Disease. ∞ indicates no upper bound in the 95% CI.
* Indicates statistical significance based on 5% type 1 error threshold.

**Figure 5:** Predictive ability of imaging and spirometry for chronic obstructive progression and all-cause mortality. Categories with fewer than 30 patients are excluded from the figure for visualization. *A, B,* Odds ratios for forced expiratory volume in 1 second ($FEV_1$) progression and, *C, D,* hazard ratios for 10-year all-cause mortality using CT-based stages only or CT-based stages and spirometry. CT-based stages are more predictive of disease progression than mortality. The combination of spirometry and CT appears to be more predictive than either alone, especially among patients with Global Initiative for Chronic Obstructive Lung Disease (GOLD) stage 2 disease with severe or very severe disease at CT.

likelihood of progression increases with the CT-based severity. The HR of mortality for patients with very severe disease (HR, 2.23 [95% CI: 1.93, 2.58]) at CT was approximately twice the HR of patients with normal-to-severe disease (severe HR, 1.11 [95% CI: 0.93, 1.32]). For the normal imaging results reference category, HRs for mild (HR, 0.83 [95% CI: 0.67, 1.04]) to severe (HR, 1.11 [95% CI: 0.93, 1.32]) disease at CT were not different.

For spirometry-based staging, only GOLD stages 1–2 showed significantly greater odds (OR, 1.53 [95% CI: 1.20, 1.95]; OR, 2.01 [95% CI: 1.55, 2.61], respectively) of $FEV_1$ progression compared with GOLD stage 0. In contrast, HRs for all-cause mortality showed a significant increasing trend with GOLD stage severity (GOLD stage 3 HR, 2.80 [95% CI: 2.38, 3.30]; GOLD stage 4 HR, 7.15 [95% CI: 6.03, 8.46]), with the exception of GOLD stage 1 (HR, 0.83 [95% CI: 0.64, 1.06]), which was not different from the GOLD stage 0 reference.
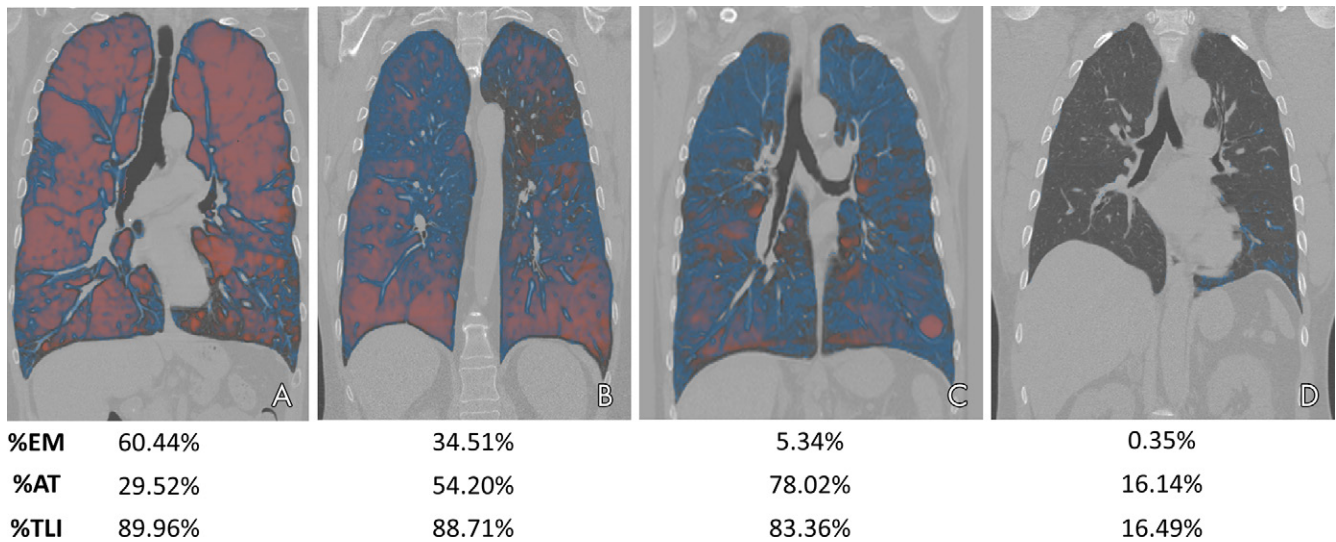
Analysis of $FEV_1$ progression using both imaging and spirometric results as predictors showed significant interactions. Among patients with GOLD stage 2 disease, severe or very severe disease at CT showed a higher likelihood of progression (4.48 [95% CI: 3.18, 6.31] and 4.72 [95% CI: 3.13, 7.13]). Mortality HRs for CT-based stages within their respective GOLD-stage categories were not significantly different from each other.

## Visualization of the Total Lung Disease Involvement Using Inspiratory Low-Attenuating Area and Attenuation Difference Maps

The spatial distribution of emphysema and air trapping across four case examples are shown using inspiratory low-attenuating areas and attenuation-difference maps in Figure 6. Inspiratory attenuations were windowed to [−1000, −900] HU, and attenuation differences were windowed to [0, 100] HU for attenuation-difference maps.

## Discussion

In this study, we leveraged deep learning and deformable registration techniques to automate quantification of emphysema and air trapping. We used these image features to define a pragmatic CT-based system for staging COPD, which parallels spirometrically defined GOLD stages. Our CT-based stages are prognostically valuable, whether used independently or in concert with pulmonary function tests, predicting both future $FEV_1$ decline and mortality. Most notably, we found that patients with moderate spirometric impairment and moderate-to-severe disease at CT had the highest likelihood of progression. These results point to the synergistic value of CT and functional testing to identify patients at highest risk of progression and to potentially identify those who might derive the greatest benefit from medical or behavioral intervention to address ongoing tobacco exposure or other inhalational exposures.

| | A | B | C | D |
|---|---|---|---|---|
| **%EM** | 60.44% | 34.51% | 5.34% | 0.35% |
| **%AT** | 29.52% | 54.20% | 78.02% | 16.14% |
| **%TLI** | 89.96% | 88.71% | 83.36% | 16.49% |

**Figure 6:** Visualization of lung involvement using attenuation difference maps and emphysema. *A,* Image in a 66-year-old man with a 35-pack-year smoking history (Global Initiative for Chronic Obstructive Lung Disease [GOLD] stage 4, forced expiratory volume in 1 second [FEV$_1$] = 23.8% predicted; FEV$_1$/forced vital capacity [FVC] ratio = 0.32). *B,* Image in a 61-year-old man with a 40-pack-year smoking history (GOLD stage 4, FEV$_1$ = 12.9% predicted, FEV$_1$/FVC ratio = 0.29). *C,* Image in a 78-year-old man with a 65.8-pack-year smoking history and chronic bronchitis (GOLD stage 2, FEV$_1$ = 59.6% predicted, FEV$_1$/FVC ratio = 0.47). *D,* Image in a 48-year-old man with a 26-pack-year smoking history (GOLD stage 0, FEV$_1$ = 92.9% predicted, FEV$_1$/FVC ratio = 0.82). %AT = percentage air trapping, %EM = percentage emphysema, %TLI = percentage total lung involvement.

Over the past several years, there has been increasing recognition of the potential prognostic value of CT for COPD. Qualitative assessments (5,9,34) and quantitative measurements (6,10,12) have each been proposed for characterization of the lung parenchyma and airways (6,10,12), with inter- and intrasoftware reproducibility of these measurements having been thoroughly assessed (35). The COPDGene investigators recently proposed a set of criteria to assist in the diagnosis of COPD, combining smoking exposure, symptoms, spirometry, and imaging to provide categories of diagnostic certainty (7)—a departure from historical definitions that did not incorporate CT imaging. This has been driven, in part, by the recognition that patients with COPD may not all exhibit abnormal pulmonary function test results. Their categories of diagnostic certainty appear to have prognostic value of their own. However, several of these measurements may be challenging to perform in the clinical environment, as they require manual labor or commercial software. We expand on these concepts, showing that open-source deep learning and deformable image registration tools can be used to automate CT-based COPD stages, which can be made readily available in a clinical picture archiving and communication system for routine use.

Similar approaches using deformable registration also have been proposed (13–16), requiring aggregation of multiple commercial software tools. Boes et al (13) and Pompe et al (14) categorically defined the proportion of lung affected by emphysema and air trapping by placing predetermined attenuation thresholds on deformably registered inspiratory and expiratory images to create parametric response maps. Ostridge et al (15) and Kirby et al (16) computed disease probability maps, describing the probability that each colocalized lung voxel contains emphysema or air trapping. In our study, quantification of air trapping is most similar to that achieved using parametric response maps,

which place thresholds on densities of deformed images. However, a fundamental difference is the removal of an upper threshold at inspiratory to define air trapping, considering voxels with an attenuation difference of less than or equal to 100 HU and inspiratory attenuation of greater than –950 HU. In an auxiliary analysis, we show that this change can improve the prediction of the GOLD stage over both the traditionally used low-attenuating area and parametric response map measures for emphysema and air trapping (Appendix E3 [supplement]). However, further investigation may be required to detail the advantages and disadvantages of each method.

Other investigators have also begun to explore the value of deep learning to automate analysis of CT. González et al (36) developed a CNN to predict the GOLD stage using CT images, and Humphries et al (20) used deep learning to automate the Fleischner visual stages of emphysema. These classification-based applications of deep learning illustrate the potential for deep learning to perform prognostication, but it may be challenging to integrate such applications into the clinical workflow or to apply quality control to them. Specifically, they are challenged by "explainability," as it can be difficult to distinguish between successful execution of the algorithm and "edge" cases at the boundaries of the capabilities of the algorithm. Algorithms that leverage deep learning to automate segmentation tasks, like those implemented in the strategy we use here, may be more readily assessed visually to ensure quality for clinical use.

Our study had several limitations. Inspiratory- and expiratory-phase imaging requires adequate patient cooperation. Suboptimal expiratory-phase images may result in apparent air trapping, leading to mischaracterization of patients. It will be important for clinical use of these algorithms to incorporate analysis of the adequacy of expiratory-phase acquisitions, which may be assessed by evaluating the concavity of the posterior

membranous portion of the trachea. Future studies may investigate the use of deep learning approaches to automatically evaluate the quality of expiratory series. Another limitation pertains to the newly recognized PRISm group of patients who have a preserved ratio but have an impaired $FEV_1pp$ of less than 80%. Studies have found that PRISm is associated with increased respiratory symptoms, inflammation, and mortality (28) and may apply to patients with overlap of fibrosis, larger body habitus, and COPD. Future studies may be helpful for evaluating the benefit of quantitative CT in the PRISm population. Finally, our proposed CT-based staging system incorporates measurements of emphysema and air trapping but does not incorporate measurements of airway caliber or wall thickening. It is possible that airway measurements may add additional prognostic value. Automated methods to standardize and accurately quantify airway abnormalities are reserved as a future direction.

In conclusion, we show that quantitative CT–based stages of COPD severity can prognosticate progression and mortality when implemented independently or in conjunction with the spirometric GOLD criteria. Fully automated deep learning algorithms can facilitate translation of the proposed stages into clinical practice, potentially allowing us to provide better care for patients with COPD.

## References

1. Centers for Disease Control and Prevention. Million hearts: strategies to reduce the prevalence of leading cardiovascular disease risk factors—United States, 2011. MMWR Morb Mortal Wkly Rep 2011;60(36):1248–1251.
2. Scanlon PD, Connett JE, Waller LA, et al. Smoking cessation and lung function in mild-to-moderate chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2000;161(2 Pt 1):381–390.
3. Vestbo J, Edwards LD, Scanlon PD, et al. Changes in forced expiratory volume in 1 second over time in COPD. N Engl J Med 2011;365(13):1184–1192.
4. Bhatt SP, Soler X, Wang X, et al. Association between functional small airway disease and FEV1 decline in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2016;194(2):178–184.
5. Lynch DA, Moore CM, Wilson C, et al. CT-based visual classification of emphysema: association with mortality in the COPDGene study. Radiology 2018;288(3):859–866.
6. Schroeder JD, McKenzie AS, Zach JA, et al. Relationships between airflow obstruction and quantitative CT measurements of emphysema, air trapping, and airways in subjects with and without chronic obstructive pulmonary disease. AJR Am J Roentgenol 2013;201(3):W460–W470.
7. 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.
8. Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 2013;187(4):347–365.
9. Lynch DA, Austin JHM, Hogg JC, et al. CT-definable subtypes of chronic obstructive pulmonary disease: a statement of the Fleischner Society. Radiology 2015;277(1):192–205.
10. Ostridge K, Wilkinson TMA. Present and future utility of computed tomography scanning in the assessment and management of COPD. Eur Respir J 2016;48(1):216–228.
11. Matsuoka S, Kurihara Y, Yagihashi K, Hoshino M, Watanabe N, Nakajima Y. Quantitative assessment of air trapping in chronic obstructive pulmonary disease using inspiratory and expiratory volumetric MDCT. AJR Am J Roentgenol 2008;190(3):762–769.
12. Lynch DA, Al-Qaisi MA. Quantitative computed tomography in chronic obstructive pulmonary disease. J Thorac Imaging 2013;28(5):284–290.
13. Boes JL, Hoff BA, Bule M, et al. Parametric response mapping monitors temporal changes on lung CT scans in the Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS). Acad Radiol 2015;22(2):186–194.
14. Pompe E, Galbán CJ, Ross BD, et al. Parametric response mapping on chest computed tomography associates with clinical and functional parameters in chronic obstructive pulmonary disease. Respir Med 2017;123:48–55.
15. Kirby M, Yin Y, Tschirren J, et al. A novel method of estimating small airway disease using inspiratory-to-expiratory computed tomography. Respiration 2017;94(4):336–345.
16. Ostridge K, Gove K, Paas KHW, et al. Using novel computed tomography analysis to describe the contribution and distribution of emphysema and small airways disease in chronic obstructive pulmonary disease. Ann Am Thorac Soc 2019;16(8):990–997.
17. Wang K, Mamidipalli A, Retson T, et al. Automated CT and MRI liver segmentation and biometry using a generalized convolutional neural network. Radiol Artif Intell 2019;1(2):180022.
18. Blansit K, Retson T, Masutani E, Bahrami N, Hsiao A. Deep learning-based prescription of cardiac MRI planes. Radiol Artif Intell 2019;1(6):e180069.
19. Bahrami N, Retson T, Blansit K, Wang K, Hsiao A. Automated selection of myocardial inversion time with a convolutional neural network: spatial temporal ensemble myocardium inversion network (STEMI-NET). Magn Reson Med 2019;81(5):3283–3291.
20. Humphries SM, Notary AM, Centeno JP, et al. Deep learning enables automatic classification of emphysema pattern at CT. Radiology 2020;294(2):434–444.
21. Gerard SE, Patton TJ, Christensen GE, Bayouth JE, Reinhardt JM. FissureNet: a deep learning approach for pulmonary fissure detection in CT images. IEEE Trans Med Imaging 2019;38(1):156–166.
22. Park B, Park H, Lee SM, Seo JB, Kim N. Lung segmentation on HRCT and volumetric CT for diffuse interstitial lung disease using deep convolutional neural networks. J Digit Imaging 2019;32(6):1019–1026.
23. Gerard SE, Herrmann J, Kaczka DW, Musch G, Fernandez-Bustamante A, Reinhardt JM. Multi-resolution convolutional neural networks for fully automated segmentation of acutely injured lungs in multiple species. Med Image Anal 2020;60(101592):101592.
24. Park J, Yun J, Kim N, et al. Fully automated lung lobe segmentation in volumetric chest CT with 3D U-Net: validation with intra- and extra-datasets. J Digit Imaging 2020;33(1):221–230.
25. National Lung Screening Trial Research Team; Aberle DR, Berg CD, et al. The National Lung Screening Trial: overview and study design. Radiology 2011;258(1):243–253.
26. Regan EA, Hokanson JE, Murphy JR, et al. Genetic Epidemiology of COPD (COPDGene) study design. COPD 2010;7(1):32–43.
27. Publications resulting from the COPDGene project. COPDGene Web site. http://www.copdgene.org/publications. Accessed January 23, 2021.
28. Wan ES, Fortis S, Regan EA, et al. Longitudinal phenotypes and mortality in preserved ratio impaired spirometry in the COPDGene study. Am J Respir Crit Care Med 2018;198(11):1397–1405.
29. Garyfallidis E, Brett M, Amirbekian B, et al. DIPY, a library for the analysis of diffusion MRI data. Front Neuroinform 2014;8:8.
30. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 2008;12(1):26–41.
31. Heussel CP, Herth FJF, Kappes J, et al. Fully automatic quantitative assessment of emphysema in computed tomography: comparison with pulmonary function testing and normal values. Eur Radiol 2009;19(10):2391–2402.

32. Madani A, Zanen J, de Maertelaer V, Gevenois PA. Pulmonary emphysema: objective quantification at multi-detector row CT—comparison with macroscopic and microscopic morphometry. Radiology 2006;238(3):1036–1043.

33. Kotikalapudi R; Keras-vis Contributors. Keras-vis. GitHub Web site. https://github.com/raghakot/keras-vis. Published 2017. Accessed May 5, 2020.

34. Kim SS, Seo JB, Lee HY, et al. Chronic obstructive pulmonary disease: lobe-based visual assessment of volumetric CT by Using standard images—comparison with quantitative CT and pulmonary function test in the COPDGene study. Radiology 2013;266(2):626–635.

35. Kirby M, Hatt C, Obuchowski N, et al. Inter- and intra-software reproducibility of computed tomography lung density measurements. Med Phys 2020;47(7):2962–2969.

36. González G, Ash SY, Vegas-Sánchez-Ferrero G, et al. Disease staging and prognosis in smokers using deep learning in chest computed tomography. Am J Respir Crit Care Med 2018;197(2):193–203.

# Erratum: Automated CT Staging of Chronic Obstructive Pulmonary Disease Severity for Predicting Disease Progression and Mortality with a Deep Learning Convolutional Neural Network