ORIGINAL ARTICLE

FEV_1/FVC Severity Stages for Chronic Obstructive Pulmonary Disease

Surya P. Bhatt[1,2], Arie Nakhmani[1,3], Spyridon Fortis[4], Matthew J. Strand[5], Edwin K. Silverman[6], Frank C. Sciurba[7]*, and Sandeep Bodduluri[1,2]*

[1]UAB Lung Imaging Lab, [2]Division of Pulmonary, Allergy and Critical Care Medicine, and [3]Department of Electrical and Computer Engineering, University of Alabama at Birmingham, Birmingham, Alabama; [4]Division of Pulmonary, Critical Care and Occupational Medicine, University of Iowa Hospital, Iowa City, Iowa; [5]Division of Biostatistics and Bioinformatics, Office of Academic Affairs, National Jewish Health, Denver, Colorado; [6]Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; and [7]Division of Pulmonary, Allergy and Critical Care Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania

ORCID IDs: 0000-0002-8418-4497 (S.P.B.); 0000-0001-7807-6740 (S.F.).

Abstract

Rationale: The diagnosis of chronic obstructive pulmonary disease (COPD) is based on a low FEV_1/FVC ratio, but the severity of COPD is classified using FEV_1% predicted ($ppFEV_1$).

Objectives: To test a new severity classification scheme for COPD using FEV_1/FVC ratio, a more robust measure of airflow obstruction than $ppFEV_1$.

Methods: In COPDGene (Genetic Epidemiology of COPD) ($N = 10{,}132$), the severity of airflow obstruction was categorized by Global Initiative for Chronic Obstructive Lung Disease (GOLD) stages 1–4 ($ppFEV_1$ of $\geq$80%, $\geq$50–80%, $\geq$30–50%, and $<$30%). A new severity classification (STaging of Airflow obstruction by Ratio; STAR) was tested in COPDGene—FEV_1/FVC $\geq$0.60 to $<$0.70, $\geq$0.50 to $<$0.60, $\geq$0.40 to $<$0.50, and $<$0.40, respectively, for stages 1–4—and applied to the combined Pittsburgh SCCOR and Emphysema COPD Research Registry for replication ($N = 2{,}017$).

Measurements and Main Results: The agreements (weighted Bangdiwala B values) between GOLD and the new FEV_1/FVC ratio severity stages were 0.89 in COPDGene and 0.88 in the Pittsburgh cohort. In COPDGene and the Pittsburgh cohort, compared with GOLD staging, STAR provided significant discrimination between the absence of airflow obstruction and stage 1 for all-cause mortality, respiratory quality of life, dyspnea, airway wall thickness, exacerbations, and lung function decline. No major differences were noted for emphysema, small airway disease, and 6-minute-walk distance. The STAR classification system identified a greater number of adults with stage 3/4 disease who would be eligible for lung transplantation and lung volume reduction procedure evaluations.

Conclusions: The new STAR severity classification scheme provides discrimination for mortality that is similar to the GOLD classification but with a more uniform gradation of disease severity. STAR differentiates patients' symptoms, disease burden, and prognosis better than the existing scheme based on $ppFEV_1$, and is less sensitive to race/ethnicity and other demographic characteristics.

Keywords: COPD, airflow obstruction, severity, staging

(Received in original form March 13, 2023; accepted in final form June 20, 2023)

*Co–senior authors.

Supported by NIH/NHLBI R01 HL151421 (S.P.B. and A.N.), UH3HL155806 (S.P.B.), NHLBI K01HL163249 (S.B.), P50 HL084948, R01 HL159805-05A1, R01 HL157879-01, NHLBI U01 HL089897, and U01 HL089856. COPDGene (Genetic Epidemiology of COPD) is also supported by the COPD Foundation through contributions made to an Industry Advisory Board that has included AstraZeneca, Bayer Pharmaceuticals, Boehringer Ingelheim, Genentech, GlaxoSmithKline, Novartis, Pfizer, and Sunovion. The funders did not have any role in the analyses or presentation of these findings.

Author Contributions: Study concept and design: S.P.B. Acquisition, analysis, or interpretation of data: all authors. Drafting of the manuscript: S.P.B. Critical revision of the manuscript for important intellectual content: all authors. Statistical analysis: S.P.B. and S.B. Study supervision: all authors.

Correspondence and requests for reprints should be addressed to Surya P. Bhatt, M.D., M.S.P.H., Division of Pulmonary, Allergy and Critical Care Medicine, University of Alabama at Birmingham, THT 422, 1720 2nd Avenue South, Birmingham, AL 35294. E-mail: sbhatt@uabmc.edu.

This article has a related editorial.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Am J Respir Crit Care Med Vol 208, Iss 6, pp 676–684, Sep 15, 2023

Originally Published in Press as DOI: 10.1164/rccm.202303-0450OC on June 20, 2023
Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the Subject: The diagnosis of chronic obstructive pulmonary disease is based on a low FEV_1/FVC ratio, but its severity is classified using the percentage predicted FEV_1, which is impacted by demographic characteristics, including race/ethnicity.

What This Study Adds to the Field: The new severity classification scheme, STaging of Airflow obstruction using Ratio, provides discrimination for mortality similar to the Global Initiative for Chronic Obstructive Lung Disease classification but with a more uniform gradation of disease severity. The STaging of Airflow obstruction using Ratio classification scheme differentiates patients' symptoms, disease burden, and prognosis better than the existing scheme based on percent predicted FEV_1 and is less sensitive to race/ethnicity and other demographic characteristics.

The diagnosis of chronic obstructive pulmonary disease (COPD) is based on spirometric airflow obstruction, and its severity is determined by gradations of reduction in percentage predicted of the FEV1 ($ppFEV_1$) as pp precedes FEV1. The clinical value of severity grading lies in its association with mortality, its causal association with respiratory symptoms, and its use in decision-making regarding certain interventions such as lung volume reduction and lung transplantation. Currently accepted recommendations for severity grading are from the Global Initiative for Chronic Obstructive Lung Disease (GOLD) and the American Thoracic Society (ATS)/European Respiratory Society, both of whom recommend specific $ppFEV_1$ thresholds or *z*-score–based cutoffs (1, 2). There are, however, several controversies in the use of $ppFEV_1$ for grading the severity of airflow obstruction. First, the percent predicted values rely on reference equations, which can be problematic when they are not representative of the general population. This has been brought into sharp focus recently with the concerns raised about corrections for race/ethnicity and the resulting potential for misdiagnosis and misclassification of severity (3–5). Second, the reference equations were derived using prebronchodilator spirometry, whereas the recommendation is to use postbronchodilator spirometry for diagnosis. Third, FEV_1 is affected by lung size; reference equations do not directly account for lung size and use height as a surrogate (6). Fourth, the presence of concomitant restriction can overestimate the severity classification when FEV_1 alone is used to make this determination. Fifth, use of the FEV_1/FVC ratio for both diagnosing the disease and determining its severity would simplify the assessment.

Severity classes should offer sufficient discrimination between categories for the prediction of clinical outcomes and for clinical decision-making. We hypothesized that using the FEV_1/FVC ratio, a more robust measure of airflow obstruction than FEV_1, for severity classification would improve associations with important outcomes such as mortality and lung function decline, computed tomography (CT)–measured emphysema and airway disease, and clinical burden including dyspnea, functional capacity, quality of life, and exacerbations. The use of FEV_1/FVC ratio to determine severity classes has been tested before, and our scheme is a modification of the classification system originally proposed by the Intermountain Thoracic Society (ITS). The ITS graded the severity of obstruction based on the FEV_1/FVC ratio in terms of confidence intervals (CIs) from the expected normal. Hegewald and colleagues included this classification scheme in their comparative study for predicting mortality and found that using $ppFEV_1$-based grades resulted in higher discrimination than FEV_1 *z*-scores and the ITS classification (7). This study was also limited by the collapsing of severity grades into mild, moderate, and severe and did not include a comparison with those with no airflow obstruction; mild airflow obstruction was the reference category (7). We derived a new severity gradation for airflow obstruction based on FEV_1/FVC ratio categories, tested our hypothesis using data from the COPDGene (Genetic Epidemiology of COPD) study, and replicated associations in the combined Pittsburgh Specialized Center of Clinically Oriented Research (SCCOR) and Emphysema COPD Research Registry cohorts.

Methods

Participants

We included participants enrolled in two large cohort studies (the multicenter COPDGene study and the combined Pittsburgh cohort). The details of these studies have been previously published (8, 9). Briefly, the COPDGene study enrolled current and former smokers aged 45–80 years. At enrollment, all participants underwent pre- and postbronchodilator spirometry using an EasyOne spirometer (ndd Medical) according to the ATS criteria. Postbronchodilator spirometry was acquired approximately 20 minutes after the administration of 180 μg of albuterol. We used postbronchodilator lung function data for all analyses. Respiratory disease–related health impairment and quality of life were assessed using the St. George's Respiratory Questionnaire (SGRQ) (10). Functional capacity was assessed using the distance walked on the 6-minute-walk test (6MWD), which was performed according to ATS guidelines. Shortness of breath was measured using the modified Medical Research Council (mMRC) dyspnea score. All participants underwent volumetric chest CT scans at maximal inspiration (i.e., total lung capacity; TLC) and at end-tidal expiration (i.e., functional residual capacity). Lung masks were applied, emphysema was quantified as the percentage of lung volume at TLC with attenuation lower than −950 HU, and air trapping was quantified as the percentage of lung volume on expiratory scans with attenuation lower than −856 HU. Parametric response mapping was used to register inspiratory and expiratory images voxel to voxel, and nonemphysematous air trapping was used to quantify functional small-airway disease (fSAD) (11). Airway wall thickness was quantified using the square root of the wall area of a theoretical airway with an internal perimeter of 10 mm (Pi10). All participants provided written informed consent before study enrollment. The COPDGene study was approved by the institutional review boards of all 21 participating centers.

The Pittsburgh subjects were enrolled from the NHLBI-funded University of Pittsburgh SCCOR and Emphysema COPD Research Center Registry cohorts (9). SCCOR includes current and former smokers aged 40–79 years with a minimum 10–pack-year tobacco history residing

around southwestern Pennsylvania. Subjects were excluded if they had other significant lung diseases, uncontrolled comorbidity including a cardiovascular event or congestive heart failure exacerbation in the preceding year, prior thoracic surgery, or a body mass index (BMI) >35 kg/m^2. Participants completed demographic and medical history questionnaires, chest CT, and spirometry (9). The Emphysema COPD Research Center Registry enrolled tobacco-exposed consenting subjects seen at the University of Pittsburgh Medical Center Comprehensive Lung Center or enrolled through the University of Pittsburgh Emphysema COPD Research Center. In the combined Pittsburgh cohort, the higher of pre- or postbronchodilator lung function data were recorded. All data-acquisition procedures were performed under a University of Pittsburgh Institutional Review Board–approved protocol, with written informed consent obtained from all participants.

To evaluate hyperinflation and air trapping between mismatched severity classes, we analyzed data from the University of Alabama at Birmingham (UAB) pulmonary function test (PFT) database. Lung volumes were acquired using nitrogen washout, and the percentage predicted TLC and residual volume (RV) were estimated. These analyses were approved by the UAB Institutional Review Board (IRB-300005811).

Follow-Up

Participants in COPDGene returned for a second visit approximately 5 years after the baseline visit, and all assessments were repeated. FEV_1 change was calculated as the annualized difference between FEV_1 at the 5-year visit and the baseline visit. Participants were contacted every 6 months by phone or using an automated telephonic system to inquire about exacerbations and vital status. Exacerbations were defined as acute worsening in symptoms requiring the use of antibiotics and/or systemic steroids. All-cause mortality was assessed on follow-up, with deaths confirmed with a combination of medical records and National Death Index.

In the combined Pittsburgh cohort, all-cause mortality was determined from the Social Security Death Index.

COPD Severity Classification

The presence of COPD was defined by postbronchodilator FEV_1/FVC ratio lower than 0.70 on spirometry according to the GOLD recommendation. The severity of airflow obstruction was categorized by GOLD stages 1–4 based on $ppFEV_1$ of ⩾80%, ⩾50–80%, ⩾30–50%, and <30%, respectively, using the Global Lung Initiative reference equations (RSpiro package) (12). A new severity classification (STaging of Airflow obstruction by Ratio; STAR) was derived using approximately the 25th, 50th, and 75th percentiles of FEV_1/FVC ratios of the participants in COPDGene with FEV_1/FVC ratio lower than 0.70. This resulted in FEV_1/FVC thresholds of ⩾0.60 to <0.70, ⩾0.50 to <0.60, ⩾0.40 to <0.50, and <0.40, respectively, for stages 1–4. The same thresholds were applied to the combined Pittsburgh cohort for replication. We calculated a modified BODE index (BMI, airflow obstruction, dyspnea, and exercise capacity) by substituting GOLD stages for the $ppFEV_1$ classes, and also derived another modified BODE index by substituting the four FEV_1/FVC ratio classes for the $ppFEV_1$ used in the BODE index (13).

Statistical Analysis

Concordance between the two severity classification schemes was assessed descriptively using Bangdiwala plots and quantified for agreement between multiple classes using the weighted Bangdiwala B value, which adjusts for the frequency of each severity class. The primary outcome for comparison of the classification schemes was all-cause mortality. Cox proportional hazards models were created with all-cause mortality as the outcome and severity classes as the predictors, with adjustments for age, sex, race, and height. Schoenfeld residuals were used to test the assumption of proportional hazards. Participants without airflow obstruction (FEV_1/FVC ⩾0.70) were treated as the reference group for all comparisons between classes for both severity classification schemes. ANOVA and Tukey's test for between-pair comparisons were used to compare differences between groups for continuous measures. Primary analyses were unadjusted for covariates because the main consideration was the effect of placing participants in discrete bins of lung function. In secondary analyses, generalized linear regression models (GLMs) were used to test associations between the severity classes and structural lung disease on CT (percent emphysema, percent fSAD, and Pi10). These models were adjusted for age, sex, race, BMI, smoking status, and pack-years of smoking. GLMs for clinical outcomes (SGRQ, mMRC, 6MWD, and FEV_1 change) were additionally adjusted for postbronchodilator FEV_1. Least-squares means derived from the GLM models were used for comparisons between stages. Zero-inflated negative binomial regression models were used to test the association between severity classes and exacerbation frequency; these models were additionally adjusted for history of exacerbations in the 1 year before enrollment. The goodness of fit of the regression models was assessed using Pearson residuals by plotting the residuals against the fitted values of the models. All analyses were performed using R statistical package v4.2.2. A two-sided α-value of 0.05 was deemed statistically significant.

Results

Participant Characteristics

We included 10,132 participants enrolled in COPDGene (after excluding 66 participants with missing spirometry) and 2,017 participants enrolled in the combined Pittsburgh cohort (after excluding 46 because of missing demographic data). The baseline characteristics of participants are shown in Table E1 in the online supplement. In COPDGene, 5,649 (56%) had no airflow obstruction and 798 (8%), 1,913 (19%), 1,167 (11%), and 605 (6%), had GOLD stages 1–4 airflow obstruction, respectively. A total of 4,668 (47.1%) were active smokers, and 5,364 (52.9%) were former smokers. In the Pittsburgh cohort, 528 (26%) had no airflow obstruction, and 213 (11%), 544 (27%), 421 (21%), and 311 (15%) had GOLD stages 1–4 airflow obstruction, respectively. A total of 654 (32.4%) were active smokers, and 1,363 (67.6%) were former smokers.

Concordance between GOLD and FEV_1/FVC Classes

The distributions of participants in COPDGene by disease severity according to GOLD and STAR classes are shown in Figure 1 (results for the Pittsburgh cohort are shown in Figure E1). The major redistributions were from GOLD stage 2 to STAR stages 1 (51.5%) and 3 (13.7%) and from GOLD stage 3 to STAR stages 2 (18.9%) and 4 (34.6%). There were smaller redistributions from GOLD stage 1 to STAR stage 2 (12.0%) and from GOLD stage 4 to STAR stage 3 (11.2%). Figure 2 shows the Bangdiwala plots for agreement between

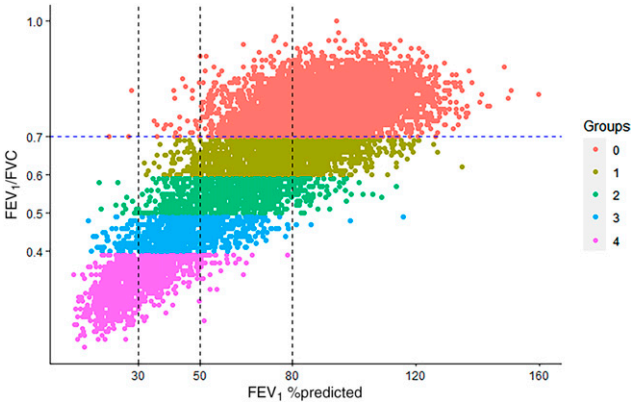
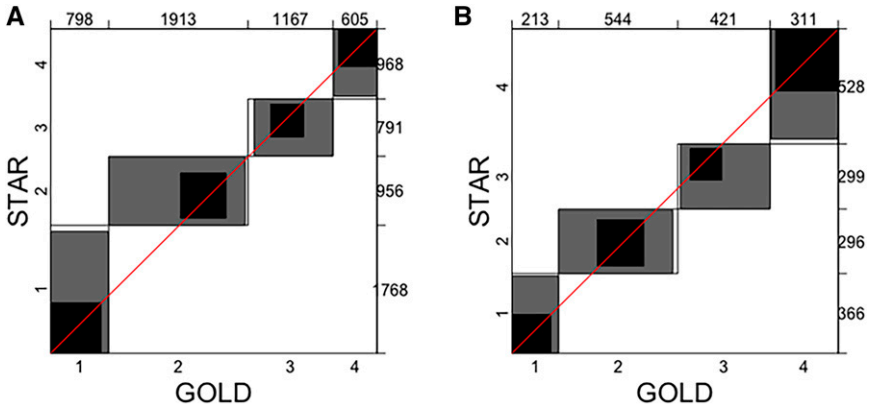


Figure 1. Distribution of disease severity by Global Initiative for Chronic Obstructive Lung Disease stage and STaging of Airflow obstruction using Ratio (STAR) severity category in the COPDGene (Genetic Epidemiology of COPD) study. Groups are STAR severity stages. COPD = chronic obstructive pulmonary disease.

GOLD and STAR severity classes. The agreements (weighted Bandiwala B values) between GOLD and STAR severity classes were 0.89 in COPDGene and 0.88 in the Pittsburgh cohort. Table E2 shows the characteristics of individuals who were reassigned stages from the GOLD scheme to the STAR scheme.

Survival

In COPDGene, over a median duration of 9.3 years (25th to 75th percentile, 4.5–10.6) (76,249 person-years), 2,200 participants died (21.7%). In the Pittsburgh cohort, over a median duration of 10.3 years (25th to 75th percentile, 5.4–13.3) (19,424 person-years), 755 participants died (37.4%). Survival curves by disease severity classes for GOLD and STAR stages are shown in Figure 3. Overall, on multivariable analyses, after adjustment for age, sex, race, and height, the discrimination for mortality was comparable between staging systems (C-statistics, 0.72 for GOLD and 0.71 for STAR). Table 1 shows unadjusted and adjusted hazard ratios for each severity stage, with no airflow obstruction as the reference group. In both COPDGene and the Pittsburgh cohort, GOLD staging did not provide any discrimination between severity stage 1 and the absence of airflow obstruction, whereas the STAR system provided a significant



C

GOLD Stage	STAR Stage	COPDGene	Pittsburgh
1	1	87.5%	84.0%
	2	12.0%	15.5%
	3	0.5%	0.5%
	4	0%	0%
2	1	51.5%	32.2%
	2	32.8%	39.5%
	3	13.7%	24.1%
	4	2.0%	4.2%
3	1	7.3%	2.9%
	2	18.9%	10.5%
	3	39.2%	34.7%
	4	34.6%	52.0%
4	1	0%	0%
	2	1.8%	1.3%
	3	11.2%	6.8%
	4	86.9%	92.0%

Figure 2. Bangdiwala agreement charts comparing classification of severity of airflow obstruction using the Global Initiative for Chronic Obstructive Lung Disease (GOLD) and STaging of Airflow obstruction using Ratio (STAR) severity scheme in (*A*) COPDGene and (*B*) Pittsburgh cohorts. (*C*) Redistribution of GOLD severity stages by STAR. The agreement plot is displayed as an $n \times n$ square, where n is the total sample size. Each large rectangle shows the maximum possible agreement given the marginal totals. Each dark rectangle indicates complete agreement. Successively lighter-shaded rectangles indicate partial agreement. Perfect agreement would be indicated by all groups showing dark perfect squares and a 45° diagonal line touching the edges of each square. The numbers on the top and right axes represent subjects placed in each GOLD and STAR stage, respectively. COPD = chronic obstructive pulmonary disease.

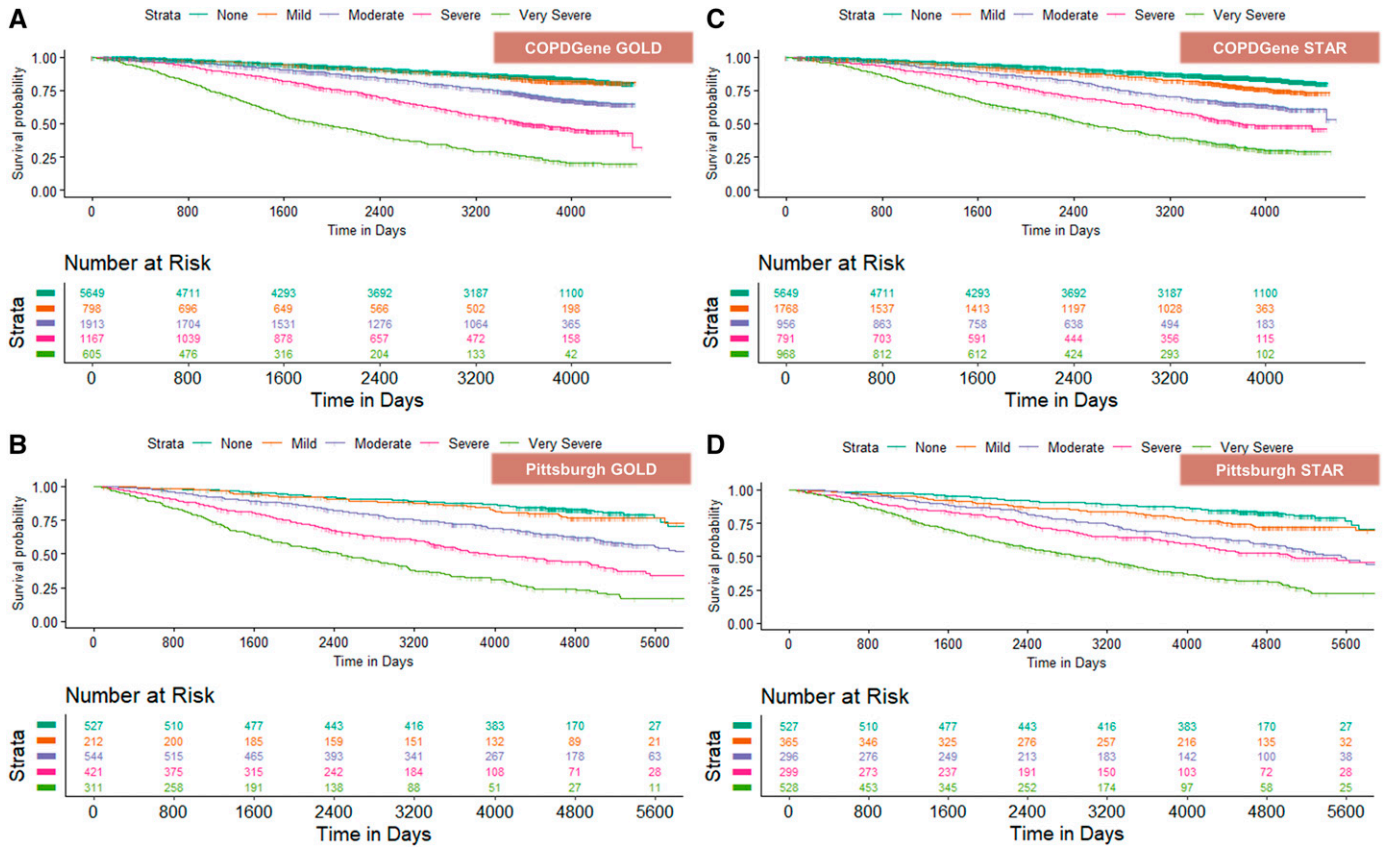


Figure 3. Kaplan-Meier curves for all-cause mortality by GOLD and STAR stages. COPDGene = Genetic Epidemiology of COPD; GOLD = Global Initiative for Chronic Obstructive Lung Disease; STAR = STaging of Airflow obstruction using Ratio.

discrimination between the absence of airflow obstruction and stage 1. STAR also provides a more uniform gradation of disease severity than GOLD. Similar results were found in analyses stratified by sex (Table E3).

Clinical Outcomes by Disease Severity

Tables 2 and E4–E6 show comparisons of the point estimates for SGRQ, mMRC dyspnea score, and 6MWD by disease severity classes. In COPDGene and the Pittsburgh cohort, GOLD staging again provided no discrimination between the absence of airflow obstruction and stage 1 for SGRQ and dyspnea, whereas significant differences were noted using the FEV_1/FVC ratio stages. The discrimination for 6MWD was variable and inconsistent across disease severity.

Structural Lung Disease by Disease Severity

Tables 3 and E7 show comparisons of the point estimates for CT-detected emphysema, CT-detected functional small-airway disease, and Pi10 by disease severity classes in COPDGene. There were significant and comparable differences by staging by either scheme for emphysema and fSAD, but there was no difference in Pi10 between GOLD stage 1 and no airflow obstruction.

Exacerbations

In COPDGene, no significant difference was noted between the absence of airflow obstruction and stage 1 by GOLD staging, in contrast to STAR staging. Compared with no airflow obstruction, after adjustment for age, sex, race, height, current smoking, pack-years of smoking, postbronchodilator FEV_1, and number of exacerbations in the previous year, the incidence rate ratios for the presence of GOLD stages 1–4 airflow obstruction were 1.15 (95% CI, 0.97–1.35; $P = 0.096$), 1.68 (95% CI, 1.51–1.86; $P < 0.001$), 2.76 (95% CI, 2.45–3.10; $P < 0.001$), and 3.32 (95% CI, 2.86–3.87; $P < 0.001$), respectively. The incidence rate ratio for STAR stages 1–4 were 1.30 (95% CI, 1.17–1.46; $P < 0.001$), 1.97 (1.74–2.24; $P < 0.001$), 2.93 (2.56–3.36; $P < 0.001$), and 3.54 (3.11–4.03; $P < 0.001$), respectively.

FEV_1 Change

In COPDGene, the annualized FEV_1 changes (least-square means) for no airflow obstruction and GOLD stages 1–4 were −32.3 (95% CI, −34.3 to −30.2), −43.4 (95% CI, −48.0 to −38.8), −50.4 (95% CI, −53.9 to −47.0), −50.0 (−55.7 to −44.4), and −41.8 (−52.4 to −31.2) ml/yr, respectively. FEV_1 changes for participants with no airflow obstruction and STAR stages 1–4 were −30.8 (95% CI, −32.8 to −28.8), −41.8 (95% CI, −44.9 to −38.6), −54.2 (95% CI, −58.9 to −49.6), −65.3 (95% CI, −71.2 to −59.4), and −61.9 (95% CI, −69.1 to −54.7) ml/yr, respectively (Figure E2).

Physiologic Considerations

We included 16,199 participants from the UAB PFT database (*see* Table E1). We compared TLC and RV for GOLD- and STAR-matched and -mismatched categories (Figures E3 and E4). We found that, within each GOLD stage, increasing STAR stages were associated with monotonic increases in hyperinflation and air trapping. In contrast, within each STAR stage, increasing GOLD

Table 1. Comparison of All-cause Mortality by Severity Stage

	COPDGene				Pittsburgh			
	Unadjusted HR (95% CI)		Adjusted HR* (95% CI)		Unadjusted HR (95% CI)		Adjusted HR* (95% CI)	
Stage	GOLD	STAR	GOLD	STAR	GOLD	STAR	GOLD	STAR
1	1.11† (0.91–1.35)	1.47 (1.29–1.68)	0.95† (0.78–1.17)	1.34 (1.66–1.53)	1.28* (0.89–1.84)	1.64 (1.22–2.19)	1.02† (0.71–1.48)	1.49 (1.11–1.99)
2	2.09 (1.86–2.35)	2.52 (2.19–2.89)	1.77 (1.57–1.99)	2.15 (1.87–2.48)	2.43 (1.89–3.13)	2.76 (2.10–3.63)	2.18 (1.69–2.80)	2.40 (1.82–3.16)
3	4.20 (3.75–4.72)	3.92 (3.44–4.47)	3.42 (3.05–3.87)	3.25 (2.84–3.73)	4.38 (3.41–5.63)	3.32 (2.53–4.35)	3.98 (3.09–5.12)	3.09 (2.35–4.06)
4	9.41 (8.32–10.64)	6.89 (6.17–7.70)	7.92 (6.98–8.98)	5.79 (5.15–6.51)	7.73 (6.02–9.94)	6.32 (4.98–8.02)	7.32 (7.67–9.44)	5.65 (4.43–7.19)

Definition of abbreviations: CI = confidence interval; COPDGene = Genetic Epidemiology of COPD; HR = hazard ratio; GOLD = Global Initiative for Chronic Obstructive Lung Disease; STAR = STaging of Airflow obstruction using Ratio.
*Model adjusted for age, sex, race, and height.
†Not statistically significant.

Table 2. Comparison of Clinical Measures by Severity Stage in the COPDGene Cohort

	No Airflow Obstruction	GOLD Stage				STAR Stage			
		1	2	3	4	1	2	3	4
No. of patients	5,649	798	1,913	1,167	605	1,768	956	791	968
SGRQ	19.8 ± 20.0	18.0 ± 17.5	33.2 ± 21.6	45.9 ± 19.1	56.1 ± 16.3	26.6 ± 21.9	36.8 ± 22.2	43.0 ± 20.2	50.9 ± 17.7
Mean Δ vs ref	Ref	−1.9	13.4	26.1	36.2	6.7	16.9	23.2	31.1
95% CI	–	−3.9 to −0.2	12.0–14.8	24.3–27.8	33.9–38.6	5.2–8.3	15.0–18.9	21.1–25.3	29.1–33.0
P value	–	0.930	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
mMRC dyspnea scale	0.9 ± 1.3	0.8 ± 1.2	1.6 ± 1.4	2.5 ± 1.2	3.1 ± 1.0	1.2 ± 1.4	1.9 ± 1.4	2.2 ± 1.3	2.9 ± 1.1
Mean Δ vs ref	Ref	−0.16	0.69	1.53	2.19	0.30	0.93	1.09	1.92
95% CI	–	−0.29 to −0.03	0.60–0.78	1.42–1.65	2.04–2.34	0.21–0.40	0.80–1.05	1.15–1.42	1.80–2.04
P value	–	0.007	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
6-min-walk distance, m	440 ± 111	461 ± 104	396 ± 111	337 ± 112	265 ± 106	421 ± 116	375 ± 119	353 ± 114	308 ± 116
Mean Δ vs ref	Ref	22	−43	−103	−175	−19	−65	−86	−131
95% CI	–	10–33	−51 to −35	−113 to −93	−188 to −162	−27 to −10	−76 to −54	−98 to −74	−142 to −120
P value	–	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Definition of abbreviations: CI = confidence interval; COPDGene = Genetic Epidemiology of COPD; GOLD = Global Initiative for Chronic Obstructive Lung Disease; mMRC = modified Medical Research Council; Ref = reference; SGRQ = St. George's Respiratory Questionnaire; STAR = STaging of Airflow obstruction using Ratio.
Values presented as mean ± SD where applicable.

Table 3. Comparison of Imaging Measures by Severity Stage in the COPDGene Cohort

	No Airflow Obstruction	GOLD Stage				STAR Stage			
		1	2	3	4	1	2	3	4
No. of patients	5,649	798	1,913	1,167	605	1,768	956	791	968
CT emphysema, %	1.9 ± 2.6	5.4 ± 5.9	7.4 ± 8.1	16.4 ± 12.6	27.1 ± 14.1	4.2 ± 4.8	8.5 ± 7.8	15.8 ± 10.9	26.7 ± 12.8
Mean Δ vs. ref	Ref	3.6	5.5	14.6	25.2	2.3	6.6	13.9	24.8
95% CI	–	2.8–4.3	5.0–6.1	13.9–15.2	24.4–26.1	1.8–2.8	6.0–7.3	13.2–14.6	24.2–25.4
P value	–	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
CT fSAD, %	7.9 ± 6.8	15.4 ± 9.6	20.6 ± 11.5	31.9 ± 11.2	37.3 ± 9.3	15.3 ± 9.7	23.9 ± 11.1	32.7 ± 10.5	36.3 ± 8.8
Mean Δ vs. ref	Ref	7.5	12.7	24.0	29.4	7.4	16.0	24.8	28.4
95% CI	–	6.5–8.5	12.0–13.4	23.1–24.9	28.2–30.6	6.7–8.1	15.1–16.9	23.9–25.8	27.5–29.3
P value	–	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Pi10, mm	2.12 ± 0.52	2.16 ± 0.47	2.61 ± 0.57	2.88 ± 0.53	2.96 ± 0.51	2.43 ± 0.60	2.68 ± 0.60	2.81 ± 0.55	2.86 ± 0.49
Mean Δ vs. ref	Ref	0.04	0.48	0.75	0.84	0.31	0.55	0.69	0.74
95% CI	–	−0.02 to 0.09	0.44–0.52	0.71–0.80	0.78–0.91	0.27–0.35	0.50–0.61	0.63–0.74	0.68–0.79
P value	–	0.323	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Definition of abbreviations: CI = confidence interval; COPDGene = Genetic Epidemiology of COPD; CT = computed tomography; GOLD = Global Initiative for Chronic Obstructive Lung Disease; fSAD = functional small-airway disease; Pi10 = square root of the wall area of a theoretical airway with an internal perimeter of 10 mm; Ref = reference; STAR = STaging of Airflow obstruction using Ratio.
Values presented as mean ± SD where applicable.

stages were associated with lower TLC and inconsistent trends for RV. Implications of STAR staging for clinical decision-making for the evaluation for lung volume reduction and lung transplantation are shown in the online supplement and in Table E8.

Discussion

We demonstrated in two large cohorts that a new severity classification scheme based on FEV_1/FVC ratio differentiates patients' symptoms, disease burden, and prognosis better than the existing scheme based on $ppFEV_1$, with a more monotonic increase in severity grades. STAR, in contrast to GOLD, provides differentiation of stage 1 from the absence of airflow obstruction. The new classification scheme based on absolute FEV_1/FVC ratio values has the added advantage of being less sensitive to race/ethnicity, which significantly impact $ppFEV_1$.

Several attempts have been made in the past to move from using $ppFEV_1$ to determine severity classes. Checkley and colleagues evaluated severity classes based on absolute FEV_1 and FEV_1/height^2 instead of $ppFEV_1$ and found that, for the prediction of several clinical outcomes, including dyspnea, quality of life, and 6MWD, these classification schemes have error rates similar to that of the GOLD classification (14). Bhatta and colleagues found that a classification based on FEV_1 standardized by the sex-specific lowest percentile (FEV_1Q) had better discrimination for mortality than $ppFEV_1$ and FEV_1 *z*-scores (15). Other studies found that $ppFEV_1$-based classification had the best discriminative accuracy for survival. In contrast to using $ppFEV_1$, a *z*-score–based method of severity classification did not result in the expected monotonic worsening of survival with increasing disease stage and had lower discrimination for survival than the GOLD classification (16). Quanjer and colleagues found minimal reclassification of subjects using a *z*-score–based classification compared with the ATS/European Respiratory Society classification based on $ppFEV_1$ (12). However, none of these studies included a comparison versus subjects without airflow obstruction; this comparison is critical for testing the validity of classifying individuals into the category of mild airflow obstruction. Severity classes should also offer sufficient discrimination between classes for predicting clinical outcomes and for clinical decision-making (17). We have extended

the literature by creating easily applicable absolute FEV_1/FVC thresholds for severity classification, and show that this classification scheme provides a more uniform gradation of disease severity than GOLD and discriminates well between the absence of and presence of mild airflow obstruction for several cross-sectional and longitudinal clinical outcomes. The STAR classification system also provides discrimination between stages for air trapping and hyperinflation, in contrast to GOLD. There is therefore a physiologic as well as a statistical case to use STAR instead of GOLD.

We also tested several additional physiologic and clinical implications of the new severity classification. Low FEV_1 has been described as a risk factor for further FEV_1 decrease and the development of worsening airflow obstruction, a phenomenon termed the horse-racing effect (18). More recent cohort studies have suggested that adults with milder airflow obstruction have a more rapid FEV_1 decrease than those with more severe disease (11, 19), a phenomenon attributed to the possibility that those with relatively preserved lung function have more to lose. However, this is equally likely to be due to a regression to the mean or nonlinearity in disease progression. Studies of structural lung disease do not support this recent finding and suggest that those with the presence of more severe disease are likely to experience faster progression (20, 21). These individuals also likely have more severe disease because they were probably on a trajectory of faster decline. We found that classifying severity using the FEV_1/FVC ratio indicates progressively greater FEV_1 decrease with worsening disease stage.

The treatment of COPD is mostly based on symptom burden and its alleviation. The severity of airflow obstruction comes into consideration when it is severe and patients may benefit from lung volume reduction or lung transplantation. We found that, even though the FEV_1/FVC ratio severity classes identify a different set of subjects as having severe or very severe disease compared with the $ppFEV_1$ classes, this does not result in any difference in the calculation of the BODE index or in the proportion of individuals who may be referred for lung transplantation evaluation based on spirometry alone. The new classification system also did not result in any reduction in the proportion of adults with COPD who may benefit from evaluation for lung volume reduction procedures, and may result in improved detection of those with greater hyperinflation and air trapping.

Given that FEV_1 is affected by FVC, using the ratio partially corrects for the variance of FEV_1 explained by FVC. We acknowledge that some individuals with air trapping may have a lower operating lung volume and hence a lower FVC. Although this can increase the FEV_1/FVC ratio and may shift the severity class to a milder stage compared with the GOLD classification, the discrimination for several clinical outcomes is not different from that of the GOLD classification, and, in some cases, may be better. Conversely, when concomitant restriction was present, Balfe and colleagues found that using the CIs of the ratio for severity classification resulted in a significant reduction in the number of subjects categorized to have severe obstruction (22). These results are in line with our results showing that, within each GOLD stage, higher STAR stages were associated with lower lung function and greater emphysema and small airway disease on CT, as well as greater hyperinflation and air trapping. The finding of lower TLC with worsening GOLD stage within each STAR stage suggests that $ppFEV_1$ likely overestimates severity in the presence of concomitant restrictive processes.

The present study has several strengths. First, we analyzed participants from two large cohorts with extensive phenotyping with clinical and imaging data, with approximately 95,693 person-years of follow-up. Second, we included individuals across a wide age range, with equal proportions of men and women and a substantial proportion of Black subjects. The discrimination for mortality was significant in analyses stratified by sex. Third, spirometry was quality-controlled, with the requirement for at least grade B according to the ATS standards for acceptability and reproducibility. Fourth, we replicated our analyses in a second cohort. Fifth, we analyzed the clinical implications of the new classification scheme. The study also has a few limitations. COPDGene included mostly current and former smokers, and hence this classification scheme needs to be validated in adults with COPD who are nonsmokers or light smokers. Although the FEV_1/FVC ratio is less sensitive than the GOLD classification to race and ethnicity, the new classification scheme needs to be validated in racial groups other than White and Black subjects.

In conclusion, we developed a new scheme to classify the severity of airflow obstruction based on FEV_1/FVC ratio, which has better discrimination for survival and better differentiates those with mild airflow obstruction from those with no airflow obstruction. By reclassifying individuals with COPD, this scheme has implications for clinical decision-making and prognostication. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

1. Agustí A, Celli BR, Criner GJ, Halpin D, Anzueto A, Barnes P, *et al.* Global Initiative for Chronic Obstructive Lung Disease 2023 Report: GOLD Executive Summary. *Am J Respir Crit Care Med* 2023;207:819–837.
2. Stanojevic S, Kaminsky DA, Miller MR, Thompson B, Aliverti A, Barjaktarevic I, *et al.* ERS/ATS technical standard on interpretive strategies for routine lung function tests. *Eur Respir J* 2022;60:2101499.
3. Baugh AD, Shiboski S, Hansel NN, Ortega V, Barjaktarevic I, Barr RG, *et al.* Reconsidering the utility of race-specific lung function prediction equations. *Am J Respir Crit Care Med* 2022;205:819–829.
4. Elmaleh-Sachs A, Balte P, Oelsner EC, Allen NB, Baugh A, Bertoni AG, *et al.* Race/ethnicity, spirometry reference equations, and prediction of incident clinical events: the Multi-Ethnic Study of Atherosclerosis (MESA) lung study. *Am J Respir Crit Care Med* 2022;205:700–710.
5. McCormack MC, Balasubramanian A, Matsui EC, Peng RD, Wise RA, Keet CA. Race, lung function, and long-term mortality in the National Health and Nutrition Examination Survey III. *Am J Respir Crit Care Med* 2022;205:723–724.
6. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, *et al.*; ERS Global Lung Function Initiative. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012;40:1324–1343.
7. Hegewald MJ, Collingridge DS, DeCato TW, Jensen RL, Morris AH. Airflow obstruction categorization methods and mortality. *Ann Am Thorac Soc* 2018;15:920–925.

8. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, *et al.* Genetic epidemiology of COPD (COPDGene) study design. *COPD* 2010;7:32–43.
9. Chandra D, Stamm JA, Palevsky PM, Leader JK, Fuhrman CR, Zhang Y, *et al.* The relationship between pulmonary emphysema and kidney function in smokers. *Chest* 2012;142:655–662.
10. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *Am Rev Respir Dis* 1992;145:1321–1327.
11. Bhatt SP, Soler X, Wang X, Murray S, Anzueto AR, Beaty TH, *et al.*; COPDGene Investigators. Association between functional small airway disease and FEV_1 decline in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2016;194:178–184.
12. Quanjer PH, Pretto JJ, Brazzale DJ, Boros PW. Grading the severity of airways obstruction: new wine in new bottles. *Eur Respir J* 2014;43: 505–512.
13. Celli BR, Cote CG, Marin JM, Casanova C, Montes de Oca M, Mendez RA, *et al.* The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med* 2004;350:1005–1012.
14. Checkley W, Foreman MG, Bhatt SP, Dransfield MT, Han M, Hanania NA, *et al.*; COPDGene Study Investigators. Differences between absolute and predicted values of forced expiratory volumes to classify ventilatory impairment in chronic obstructive pulmonary disease. *Respir Med* 2016; 111:30–38.
15. Bhatta L, Leivseth L, Mai XM, Henriksen AH, Carslake D, Chen Y, *et al.* Spirometric classifications of chronic obstructive pulmonary disease severity as predictive markers for clinical outcomes: the HUNT Study. *Am J Respir Crit Care Med* 2021;203:1033–1037.
16. Tejero E, Prats E, Casitas R, Galera R, Pardo P, Gavilán A, *et al.* Classification of airflow limitation based on z-score underestimates mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2017;196:298–305.
17. Fletcher C, Peto R. The natural history of chronic airflow obstruction. *BMJ* 1977;1:1645–1648.
18. Peto R. The horse-racing effect. *Lancet* 1981;2:467–468.
19. Kim J, Yoon HI, Oh YM, Lim SY, Lee JH, Kim TH, *et al.* Lung function decline rates according to GOLD group in patients with chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2015; 10:1819–1827.
20. Bhatt SP, Bodduluri S, Hoffman EA, Newell JD Jr, Sieren JC, Dransfield MT, *et al*; COPDGene Investigators. CT measure of lung at-risk and lung function decline in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2017;196:569–576.
21. Bhatt SP, Bodduluri S, Dransfield MT, Reinhardt JM, Crapo JD, Silverman EK, *et al.* Acute exacerbations are associated with progression of emphysema. *Ann Am Thorac Soc* 2022;19:2108–2111.
22. Balfe DL, Lewis M, Mohsenifar Z. Grading the severity of obstruction in the presence of a restrictive ventilatory defect. *Chest* 2002;122: 1365–1369.