



Published in final edited form as:

*Respir Med.* 2016 February ; 111: 30–38. doi:10.1016/j.rmed.2015.11.004.

## Differences between absolute and predicted values of forced expiratory volumes to classify ventilatory impairment in chronic obstructive pulmonary disease

William Checkley<sup>a,b,\*</sup>, Marilyn G. Foreman<sup>c</sup>, Surya P. Bhatt<sup>d</sup>, Mark T. Dransfield<sup>d</sup>, MeiLan Han<sup>e</sup>, Nicola A. Hanania<sup>f</sup>, Nadia N. Hansel<sup>a</sup>, Elizabeth A. Regan<sup>g</sup>, and Robert A. Wise<sup>a</sup>  
**COPDGene Study Investigators**

<sup>a</sup>Division of Pulmonary and Critical Care, School of Medicine, Johns Hopkins University, Baltimore, USA

<sup>b</sup>Program in Global Disease Epidemiology and Control, Department of International Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, USA

<sup>c</sup>Division of Pulmonary and Critical Care Medicine, Morehouse School of Medicine, Atlanta, USA

<sup>d</sup>Division of Pulmonary, Allergy and Critical Care Medicine, School of Medicine, University of Alabama at Birmingham, Birmingham, USA

<sup>e</sup>Division of Pulmonary and Critical Care Medicine, School of Medicine, University of Michigan, Ann Arbor, USA

<sup>f</sup>Section of Pulmonary, Critical Care, and Sleep Medicine, Baylor College of Medicine, Houston, USA

<sup>g</sup>Division of Rheumatology, Department of Medicine, National Jewish Health, Denver, USA

### Abstract

The Global Initiative for Chronic Obstructive Lung Disease (GOLD) severity criterion for COPD is used widely in clinical and research settings; however, it requires the use of ethnic- or population-specific reference equations. We propose two alternative severity criteria based on absolute post-bronchodilator FEV<sub>1</sub> values (FEV<sub>1</sub> and FEV<sub>1</sub>/height<sup>2</sup>) that do not depend on reference equations. We compared the accuracy of these classification schemes to those based on % predicted values (GOLD criterion) and Z-scores of post-bronchodilator FEV<sub>1</sub> to predict COPD-related functional outcomes or percent emphysema by computerized tomography of the lung. We tested the predictive accuracy of all severity criteria for the 6-minute walk distance (6MWD), St. George's Respiratory Questionnaire (SGRQ), 36-item Short-Form Health Survey physical health

\*Corresponding author. Division of Pulmonary and Critical Care, School of Medicine, Johns Hopkins University, 1800 Orleans Ave Suite 9121, Baltimore, USA. wcheckl1@jhmi.edu (W. Checkley).

#### Contributorship

WC and RW developed the idea for this manuscript and wrote the first draft. WC led the statistical analysis. MF, SB, MD, MH, NAH, NNH, ER, RW participated in the writing of manuscript, provided important intellectual content and gave their final approval of the version submitted for publication.

#### Conflicts of interest

The authors report no conflicts of interest.

component score (SF-36) and the MMRC Dyspnea Score. We used 10-fold cross-validation to estimate average prediction errors and Bonferroni-adjusted t-tests to compare average prediction errors across classification criteria. We analyzed data of 3772 participants with COPD (average age 63 years, 54% male). Severity criteria based on absolute post-bronchodilator FEV<sub>1</sub> or FEV<sub>1</sub>/height<sup>2</sup> yielded similar prediction errors for 6MWD, SGRQ, SF-36 physical health component score, and the MMRC Dyspnea Score when compared to the GOLD criterion (all  $p > 0.34$ ); and, had similar predictive accuracy when compared with the Z-scores criterion, with the exception for 6MWD where post-bronchodilator FEV<sub>1</sub> appeared to perform slightly better than Z-scores ( $p = 0.01$ ). Subgroup analyses did not identify differences across severity criteria by race, sex, or age between absolute values and the GOLD criterion or one based on Z-scores. Severity criteria for COPD based on absolute values of post-bronchodilator FEV<sub>1</sub> performed equally as well as did criteria based on predicted values when benchmarked against COPD-related functional and structural outcomes, are simple to use, and may provide a more accessible and comparable approach to severity classification worldwide, especially in settings where prediction equations are not available.

## Keywords

COPD; Reference equations; Severity criteria

---

## 1. Introduction

Several methods have been used to classify severity of ventilatory impairment (i.e., airflow obstruction) in chronic obstructive pulmonary disease (COPD). Currently, the most widely adopted standard to classify ventilatory impairment in COPD is based on the Global Initiative for Chronic Obstructive Lung Disease (GOLD) classification which stratifies patients by percent predicted FEV<sub>1</sub> [1]. Similarly, the joint American Thoracic Society and European Respiratory Society guidelines also rank severity of obstructive ventilatory defects using percent predicted forced expiratory values, but use different stratification [2]. These classification schemes are based on the percent predicted FEV<sub>1</sub>, which, in turn, depends upon the population being studied and the reference equations that are being used.

The use of reference equations to classify disease severity may lead to difficulties in applying and interpreting this information in both individuals and research populations. Because reference equations differ between laboratories and research studies, people with the same lung function may be categorized differently depending on which reference equations were used. The use of reference equations for stratifying ventilatory impairment may also lead to some confusing conclusions regarding functional impairment. For example, the GOLD staging system using reference equations derived for Caucasians in NHANES would classify a 68-inch tall 75 year-old man with COPD and an FEV<sub>1</sub> of 1.5 L to have moderate disease (Stage II, 53% predicted) whereas a 40 year-old man with COPD of the same height and the same FEV<sub>1</sub> would be classified as having severe disease (Stage III, 38% predicted). Since maximum ventilatory capacity with exercise is determined by multiplying the FEV<sub>1</sub> by 35 [3], we would expect that functional capacity would be similar for a ventilatory-limited individual with the same absolute FEV<sub>1</sub> regardless of age, sex or race. In

the above example, the older individual would have similar or greater ventilatory impairment and more functional limitation compared to the younger individual even though the former would be classified as having more severe disease.

Furthermore, using percent predicted to stratify severity of ventilatory impairment may lead to inconsistent assessments of severity across races and sex. Because predicted values of lung function are lower for women and African Americans, compared to men and Caucasians, women and African-Americans might need to have lower lung function to qualify for workers compensation or disability benefits when their degree of impairment is equivalent. Similarly, reference equations derived from resource-poor settings in low- and middle-income countries might underestimate the magnitude of ventilatory impairment in epidemiologic studies [4].

Miller and Pedersen examined spirometric predictors of mortality in a large general population sample, and found that multiples of absolute FEV<sub>1</sub> representing the lowest 1 percentile of the population were a stronger predictor of survival than percent predicted FEV<sub>1</sub> [5]. They also found that absolute FEV<sub>1</sub> values divided by height-cubed or height-squared provided better prediction of survival in a general population and in a COPD study sample than did FEV<sub>1</sub> alone [6,7]. In our proposed approach, we modified the Miller–Pedersen model to determine whether it can be used to classify severity of functional or structural impairment in a COPD population. We sought to test the hypothesis that a simplified classification system for severity of obstructive ventilatory defects based on either absolute FEV<sub>1</sub>, or height-adjusted FEV<sub>1</sub> irrespective of age, sex, or race would be predictive of functional limitations, disease impact, quality of life, and severity of emphysema. Moreover, we hypothesized that use of absolute values for the classification of ventilatory impairment would perform similarly in the statistical prediction of COPD-related functional outcomes as would the GOLD classification or one based on Z-scores. We also wanted to know not only about functional characteristics, but also the structural or anatomic measures of COPD. To test these hypotheses, we analyzed data from the COPDGene study which included subjects with a wide range of lung function abnormalities and measures of functional impairment, disease impact, quality of life, and percent emphysema by computerized tomography of the lung.

## 2. Methods

### 2.1. Study setting

The COPDGene study is a multicenter investigation of the genetic epidemiology of smoking-related lung disease which recruited 10,300 subjects at 21 clinical centers. Subjects were selected for participation if they were: aged 45–80 years; smoked cigarettes for 10 pack-years; and, were willing to undergo testing that included spirometry, chest CT scan, and blood collection for biomarker and genetic analysis [8]. Participants in the COPDGene Study were two-thirds non-Hispanic white and one-third black. All participants provided written informed consent. The study was approved by the Institutional Review Boards at participating institutions.

## 2.2. Study design

We selected participants from COPDGene who had post-bronchodilator spirometry data available for analysis and identified those who met criteria for COPD. We defined COPD as a post-bronchodilator  $FEV_1/FVC < \text{lower limit of normal (LLN)}$  using the Global Lung Function Initiative reference equations that adjust for age, sex, height, and race [9]. We defined severity of ventilatory impairment using four different criteria (Table 1). The first was based on % predicted values of  $FEV_1$  [1] using the Global Lung Function Initiative reference equations that adjust for age, sex, height, and race [9]. The second and third were based on absolute values of  $FEV_1$  and height-adjusted  $FEV_1$ , respectively. The last criterion was based on post-bronchodilator  $FEV_1$  Z-score thresholds [10], using the Global Lung Function Initiative reference equations that adjust for age, sex, height, and race [9]. We then selected *a priori* four COPD-related functional outcomes for evaluation: 6-minute walk distance (6MWD), St. George's Respiratory Questionnaire (SGRQ), 36-item Short-Form Health Survey (SF-36) physical health component score and the Modified Medical Research Council (MMRC) Dyspnea Score. We also examined percent emphysema on chest CT scan as the COPD-related structural outcome. Chest CT scans were performed using multi-detector helical CT scanners with 16 or more detectors. Protocols for scanner types have been published [11]. Emphysema and airway disease severity and distribution were obtained from the inspiratory CT acquisition and air trapping from the expiratory acquisition. SLICER software ([www.slicer.org](http://www.slicer.org)) was used to calculate percent emphysema and air trapping and defined as percent of lung voxels less than  $-950$  and  $-856$  Hounsfield units, respectively.

## 2.3. Biostatistical methods

The primary objective was to compare the ability of different ventilatory impairment staging criteria to predict COPD-related functional outcomes. We used ten-fold cross validation to measure prediction error among these criteria. Specifically, we used the root mean square error (RMSE) to measure the expected prediction error [12]. To perform this analysis, we first separated participants with COPD into ten approximately equal subsets. Each subset consisted of a simple random sample without replacement. We then used regression models to perform a statistical prediction for all combinations of each outcome as the dependent variable and categories of each severity criteria as the independent variable. We conducted ten regressions excluding one of the ten subsets during each run and performed a statistical prediction for each outcome. We then used Bonferroni-adjusted t-tests to compare the RMSE between absolute and predicted severity staging systems. We included indicator variables for site as fixed effects to account for heterogeneity in COPD-related functional outcomes. We used linear regression for three of the functional outcomes (6MWT, SF-36 physical health component score and SGRQ) and ordinal logistic regression for MMRC Dyspnea Score Scale. We also conducted subgroup analyses as above by race (white or black), sex and age ( $<65$  years or  $\geq 65$  years) to identify subgroups in which severity criteria may perform differently. As a secondary analysis, we graphically examined the relationship between percent emphysema and severity according to the three types of severity criteria. We conducted all statistical analyses in R ([www.r-project.org](http://www.r-project.org)).

### 3. Results

#### 3.1. Participant characteristics

A total of 10,237 participants had spirometry data available for analysis, of which 3772 (37%) met criteria for COPD. Participants with COPD were older (average age 62.6 vs. 57.9 years;  $p < 0.001$ ), were equally likely to be male (54% vs. 53%;  $p = 0.18$ ), were more likely to be white (76% vs. 62%;  $p < 0.001$ ), weighed more (average BMI 29.4 vs. 27.7 kg/m<sup>2</sup>;  $p < 0.001$ ), were heavier smokers (average pack-years 51.7 vs. 39.8;  $p < 0.001$ ) and had a lower resting oxygen saturation (average 94.8% vs. 96.9%;  $p < 0.001$ ) than did those without COPD at enrollment. We also selected participant characteristics by each of the three severity classification criteria in Table 2. Differences were apparent in age, sex, and BODE index when stratified by severity categories across the three classification criteria. For example, there was a higher proportion of males than in the mild category for absolute FEV<sub>1</sub> and FEV<sub>1</sub>/height<sup>2</sup> than for % predicted FEV<sub>1</sub>. In contrast, there was a lower proportion of males the mild category for absolute FEV<sub>1</sub> than for % predicted FEV<sub>1</sub> or FEV<sub>1</sub>/height<sup>2</sup>.

#### 3.2. COPD-related outcomes for absolute vs. percent predicted values of FEV<sub>1</sub>

In Table 3, we present summary statistics of the four COPD-related outcomes stratified by severity for each of the three approaches to classify ventilatory impairment. On visual inspection, the two proposed approaches to classify ventilatory impairment based on absolute post-bronchodilator FEV<sub>1</sub> values had similar predictive errors as did the GOLD severity staging system or Z-scores (Fig. 1). Under formal evaluation, the approaches based on absolute post-bronchodilator FEV<sub>1</sub> values (FEV<sub>1</sub> or FEV<sub>1</sub>/height<sup>2</sup>) had RMSEs that were not different from that of the GOLD criterion (Table 2). In addition, the approaches based on absolute post-bronchodilator FEV<sub>1</sub> values (FEV<sub>1</sub> or FEV<sub>1</sub>/height<sup>2</sup>) had RMSEs that were not different from that based on Z-scores (all  $p \geq 0.06$ ), but for one exception: absolute post-bronchodilator FEV<sub>1</sub> values appeared to perform slightly better than Z-scores when predicting 6WMD outcomes ( $p = 0.01$ ). The distribution of emphysema based on CT scan findings increased inversely with severity in similar fashion across all classification criteria (Fig. 2).

#### 3.3. Subgroup analyses

We did not identify important differences in COPD-related outcomes between the GOLD severity criterion and the two severity criteria based on absolute values for either whites (all  $p \geq 0.31$ ) or blacks (all  $p = 1$ ); for either men (all  $p = 1$ ) or women (all  $p \geq 0.20$ ); and, for either adults aged <65 years (all  $p \geq 0.54$ ) or those aged  $\geq 65$  years (all  $p \geq 0.94$ ; Figs. 3–5). We also did not find differences between Z-score severity criterion for blacks, men or women, and, adults aged <65 years or those aged  $\geq 65$  years (all  $p > 0.09$ ; Figs. 3–5). The only exception that the absolute FEV<sub>1</sub> value had better predictive accuracy for the 6MWD than did the Z-score criterion ( $p = 0.008$ ), but a worse predictive accuracy for the SGRQ ( $p = 0.008$ ).

## 4. Discussion

In this analysis of a large number of well-characterized patients with COPD, we compared four different methods of grading severity of airflow obstruction with respect to their ability to statistically predict functional impairment. We also graphically contrasted differences in structural impairment. The main finding of this analysis is that absolute values of post-bronchodilator FEV<sub>1</sub> were equally effective in predicting functional impairment and anatomic abnormalities as did severity ranking based on percent predicted or Z-scores of post-bronchodilator FEV<sub>1</sub>. Subgroup analyses did not identify major differences across severity criteria by race (white or black), sex, or age (<65 or ≥65 years).

The rationale for use of absolute measures of post-bronchodilator FEV<sub>1</sub> to classify ventilatory impairment is based on evidence from Miller and Pedersen [6] who found that multiples of the lowest 1 percentiles of FEV<sub>1</sub> in the general population was superior to the use of FEV<sub>1</sub> percent predicted in predicting mortality. In their study, they used multiples of 0.5 L in men and 0.4 L in women as a prognostic indicator. We simplified their approach by using multiples of 0.5 L in both men and women and deriving a stratification scheme that corresponded to the descriptors used for the GOLD classification of ventilatory impairment.

In general, the four classification criteria for ventilatory impairment performed similarly; however, we observed some subtle differences. The use of absolute levels of FEV<sub>1</sub> classified fewer people as having very severe impairment compared to the GOLD criteria, whereas the use of height-adjusted FEV<sub>1</sub> put more people into a very severe category, but distributed the population more evenly across categories. Absolute FEV<sub>1</sub> performed less well in predicting the SGRQ total score than percent predicted. This might reflect that younger people experience a greater impact of disease on activities and symptoms than do older individuals. In terms of the median MMRC Dyspnea Scale, use of absolute FEV<sub>1</sub> gave a better distribution of severity than did the other methods.

Although all four classification criteria for ventilatory impairment performed equally well in the population under study when predicting COPD-related functional outcomes, we think that there are some advantages to using measures of absolute FEV<sub>1</sub> to classify severity of ventilatory limitation in COPD rather than the GOLD or ATS/ERS criteria that are based on percent predicted, or criteria based on Z-scores. First, the simplicity of the classification permits it to be memorized and applied easily at the bedside without the need to refer to reference equations. Second, it avoids the confusion that may attend classification of an individual who might be tested at different laboratories using different reference equations that give different grades of severity for the same FEV<sub>1</sub>. Third, our proposed approach may be used to standardize classification of functional ventilatory limitation across age ranges, gender, races, and geographic location without having to use different reference equations. Finally, although not tested in our analysis, an approach based on absolute values of FEV<sub>1</sub> has been found to be superior to use of percent predicted FEV<sub>1</sub> in predicting mortality [6].

Despite these advantages, we do see that there can be some disadvantages to using either of our proposed approaches when compared to the GOLD severity system. The GOLD staging system has been widely used in COPD clinical trials and population studies to describe and

stratify the population and there may not be a direct correspondence between cross-sectional measures of functional limitation as we have used in this study and longitudinal measures of disease worsening or response to treatment. Moreover, in the small proportion of the population who are at extremes of height, the use of absolute measures of FEV<sub>1</sub> rather than percent predicted or height-adjusted FEV<sub>1</sub> may give misleading inferences. In addition, the use of percent predicted values implies that there has been a loss of lung function compared to the reference population so this might be better suited for assessment of impairment in proceedings like toxic torts or worker compensation judgments. The use of percent predicted in the assessment of severity of ventilatory limitation is so widely embedded in the thinking and practice of pulmonary medicine that there would be a natural resistance to change without a stronger rationale than we present here. Finally, although absolute values of post-bronchodilator FEV<sub>1</sub> were equally effective in predicting functional impairment and anatomic abnormalities as severity ranking based on percent predicted or Z-scores of post-bronchodilator FEV<sub>1</sub> when cross-sectional data are used, the results may be different when longitudinal data are used.

This study has several obvious strengths. It follows up an approach to assessment of mortality in a large European general population sample using a population of patients with lung disease [5]. We used a large cohort of well-characterized patients at multiple sites across the United States [8]. We had multiple measures of functional impairment including six-minute walk distance, MMRC dyspnea scale, disease specific and general measures of quality of life (SGRQ and SF-36 physical health component scores, respectively). In addition, we were also able to compare the different approaches for prediction of anatomic destruction of the lung assessed by computed tomography. Our analysis also has some shortcomings. We do not have longitudinal outcome data including survival in this population to assess whether one schema is better than another to predict progression or mortality. We cannot say whether the assessments of ventilatory impairment in this population can also be applied to a general population sample or patients with restrictive or mixed restrictive and obstructive disease. Finally, the COPDGene study was limited to white and blacks, and thus our findings may not be generalizable to other races.

In summary, our analyses suggest that a simple classification of ventilatory impairment based on height-adjusted absolute FEV<sub>1</sub> values may be as effective as traditional schema using percent predicted FEV<sub>1</sub>. Furthermore, the former is easy to apply and does not rely on selection and calculation of reference values, and may provide a better assessment of prognosis than percent predicted.

## Acknowledgments

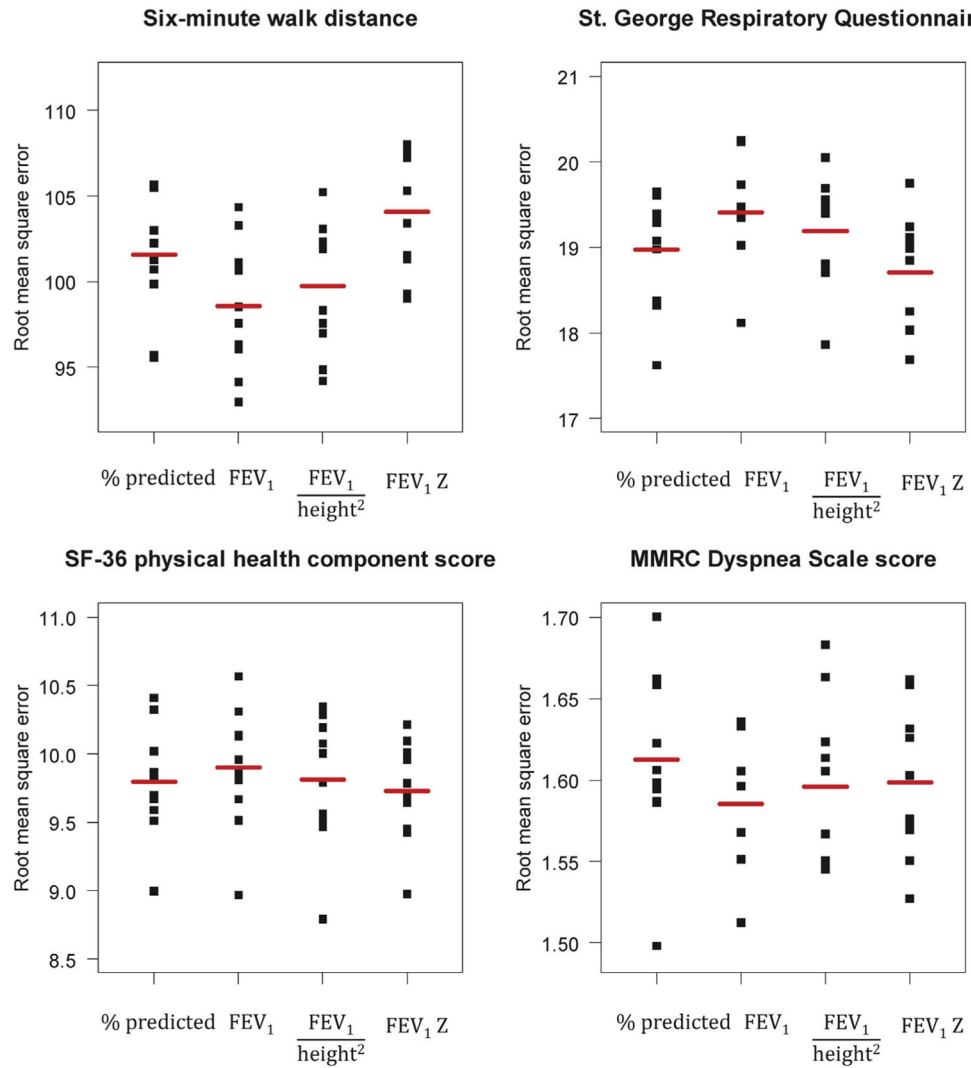
### Funding

COPDGene was supported by Award Numbers U01HL089897 and U01HL089856 from the National Heart, Lung, and Blood Institute. William Checkley was further supported by a Pathway to Independence Award (R00HL096955) from the National Heart, Lung and Blood Institute.

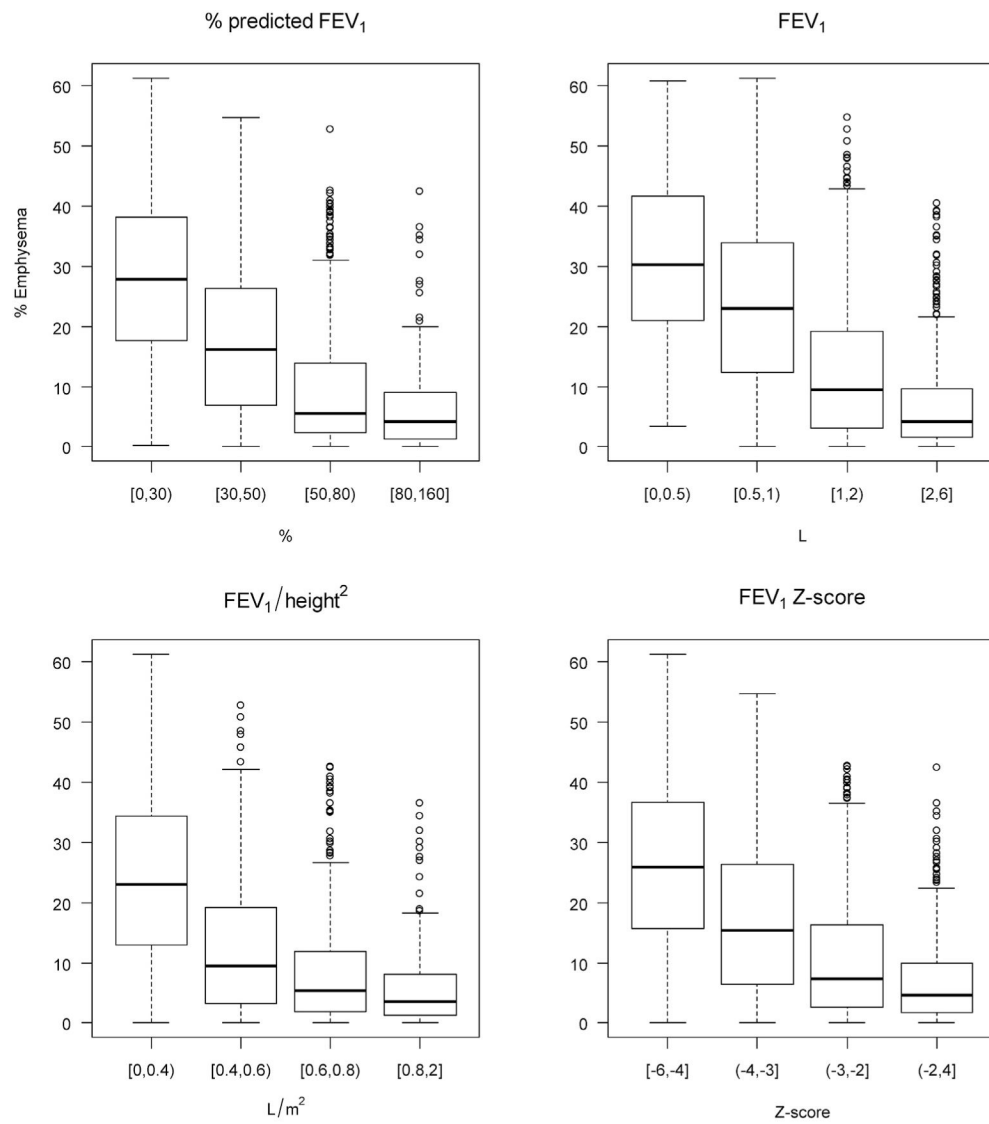
## References

1. Vestbo J, Hurd S, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, Barnes PJ, Fabbri LM, Martinez FJ, Nishimura M, Stockley RA, Sin DD, Rodriguez-Roisin R. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2013; 187:347–365. [PubMed: 22878278]
2. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, Coates A, van der Grinten CP, Gustafsson P, Hankinson J, Jensen R, Johnson DC, MacIntyre N, McKay R, Miller MR, Navajas D, Pedersen OF, Wanger J. Interpretative strategies for lung function tests. *Eur Respir J*. 2005; 26:948–968. [PubMed: 16264058]
3. Gandevia B, Hugh-Jones P. Terminology for measurements of ventilatory capacity; a report to the thoracic society. *Thorax*. 1957; 12:290–293. [PubMed: 13496030]
4. Duong M, Islam S, Rangarajan S, Teo K, O'Byrne PM, Schünemann HJ, Igumbor E, Chifamba J, Liu L, Li W, Ismail T, Shankar K, Shahid M, Vijayakumar K, Yusuf R, Zatonka K, Oguz A, Rosengren A, Heidari H, Almahmeed W, Diaz R, Oliveira G, Lopez-Jaramillo P, Seron P, Killian K, Yusuf S. PURE-BREATH Study Investigators. Global differences in lung function by region (PURE): an international, community-based prospective study. *Lancet Respir Med*. 2013; 1:599–609. [PubMed: 24461663]
5. Miller MR, Pedersen OF. New concepts for expressing forced expiratory volume in 1 s arising from survival analysis. *Eur Respir J*. 2010; 35:873–882. [PubMed: 19741033]
6. Miller MR, Pedersen OF, Lange P, Vestbo J. Improved survival prediction from lung function data in a large population sample. *Respir Med*. 2009; 103:442–448. [PubMed: 18993043]
7. Miller MR, Pedersen OF, Dirksen A. Improved staging of chronic obstructive lung disease. *Int J COPD*. 2007; 2:657–663.
8. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, Curran-Everett D, Silverman EK, Crapo JD. Genetic epidemiology of COPD (COPDGene) study design. *COPD*. 2010; 7:32–43. [PubMed: 20214461]
9. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, Enright PL, Hankinson JL, Ip MS, Zheng J, Stocks J. 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. [PubMed: 22743675]
10. 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. [PubMed: 23988764]
11. Han MK, Kazerooni EA, Lynch DA, Liu LX, Murray S, Curtis JL, Criner GJ, Kim V, Bowler RP, Hanania NA, Anzueto AR, Make BJ, Hokanson JE, Crapo JD, Silverman EK, Martinez FJ, Washko GR. COPDGene Investigators. Chronic obstructive pulmonary disease exacerbations in the COPDGene study: associated radiologic phenotypes. *Radiology*. 2011; 261:274–282. [PubMed: 21788524]
12. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*. Springer-Verlag; New York: 2001. p. 18-21.p. 214

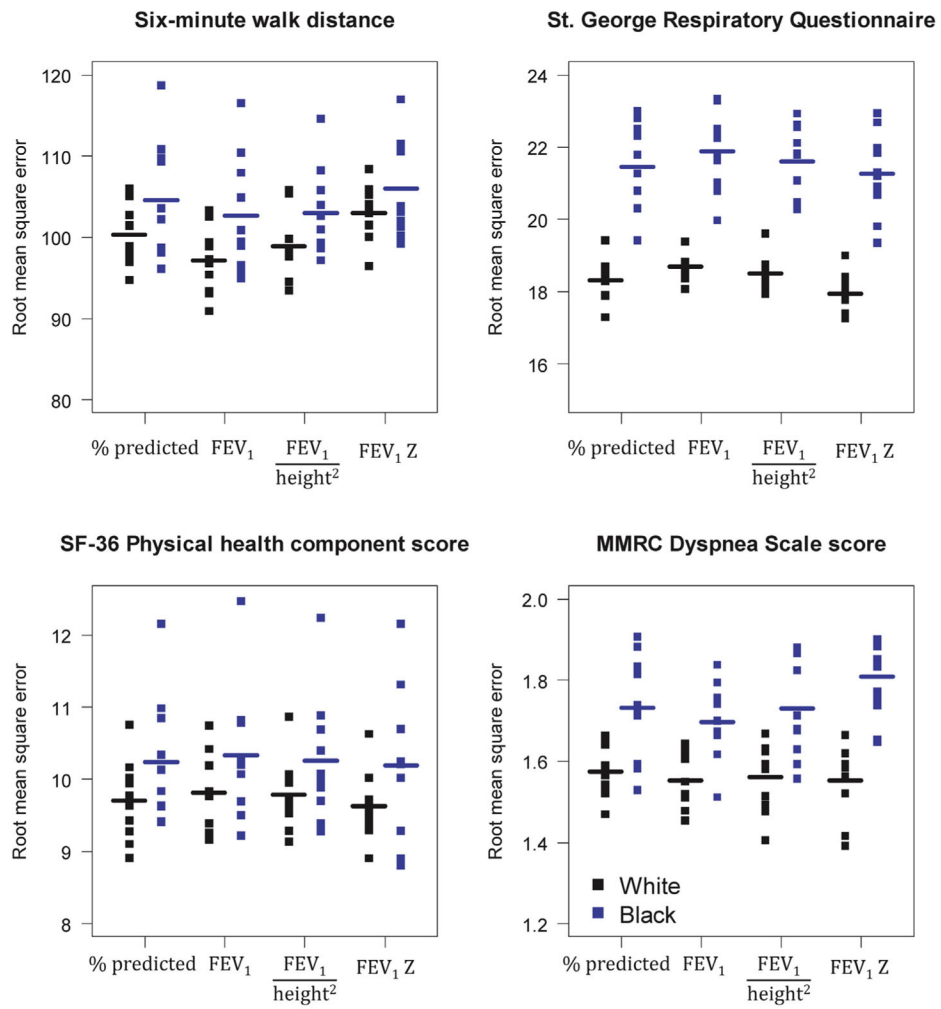




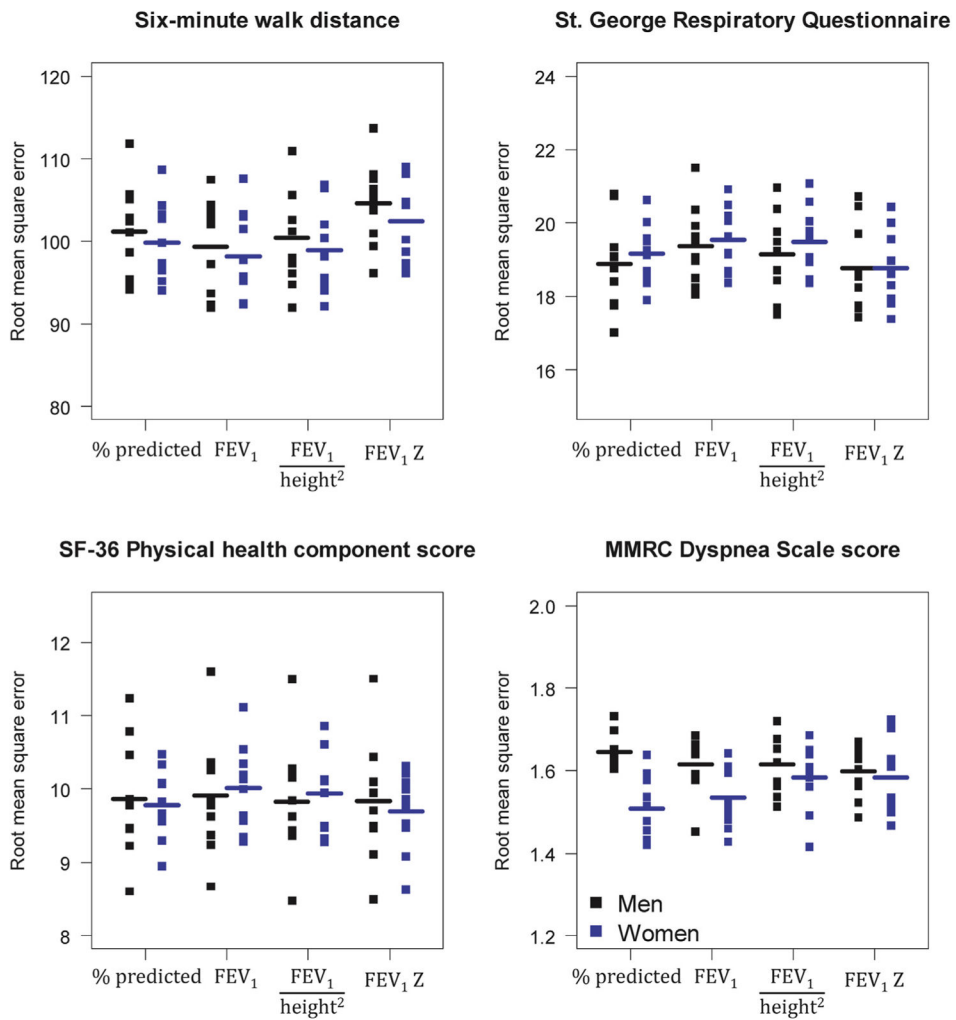
**Fig. 1.** Distribution of the 10-fold square root expected prediction errors for six-minute walk distance, St. George Respiratory Symptoms Questionnaire Score, SF-36 Physical health component score, and the MMRC Dyspnea Scale Score by severity classification criteria for ventilatory impairment. The red horizontal lines represent the root mean squared errors.



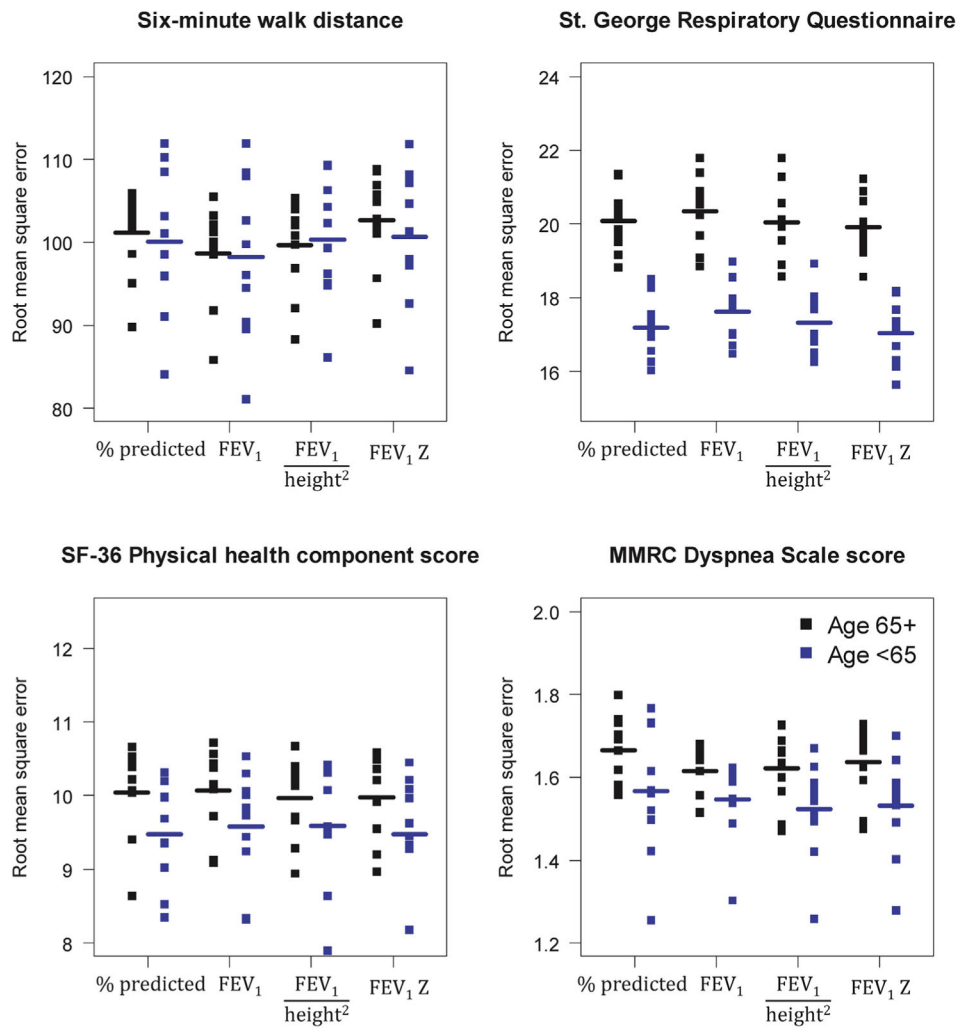
**Fig. 2.** Boxplots of % emphysema stratified by the three different severity classification criteria for ventilatory impairment.



**Fig. 3.** Distribution of the 10-fold square root expected prediction errors for six-minute walk distance, St. George Respiratory Symptoms Questionnaire Score, SF-36 physical health component score, and the MMRC Dyspnea Scale Score by severity classification criteria for ventilatory impairment stratified by race. The red horizontal lines represent the root mean squared errors.



**Fig. 4.** Distribution of the 10-fold square root expected prediction errors for six-minute walk distance, St. George Respiratory Symptoms Questionnaire Score, SF-36 physical health component score, and the MMRC Dyspnea Scale Score by severity classification criteria for ventilatory impairment stratified by sex. The red horizontal lines represent the root mean squared errors.



**Fig. 5.** Distribution of the 10-fold square root expected prediction errors for six-minute walk distance, St. George Respiratory Symptoms Questionnaire Score, SF-36 physical health component score, and the MMRC Dyspnea Scale Score by severity classification criteria for ventilatory impairment stratified by age. The red horizontal lines represent the root mean squared errors.

**Table 1**

Approaches to classify ventilatory impairment.

	<u>Proposed approaches to classify ventilatory impairment</u>		<u>Z-scores to classify ventilatory impairment</u>	<u>GOLD guidelines to classify ventilatory impairment</u>
	<b>FEV<sub>1</sub> (L)</b>	<b>FEV<sub>1</sub>/height<sup>2</sup> (L/m<sup>2</sup>)</b>	<b>FEV<sub>1</sub> Z-score</b>	<b>% Predicted FEV<sub>1</sub></b>
Mild	2	0.8	>-2	80
Moderate	1 to 1.99	0.6 to 0.79	-2.99 to -2.0	50 to 79
Severe	0.5 to 0.99	0.4 to 0.59	-3.99 to -3.0	30 to 49
Very Severe	0 to 0.49	0 to 0.39	-4	0 to 29

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Selected participant characteristics stratified by severity for the three different classification criteria of ventilatory impairment.

	Mild				Moderate				Severe				Very severe			
	% predicted FEV <sub>1</sub>	FEV <sub>1</sub> (L)	FEV <sub>1</sub> /height <sup>2</sup> (L/m <sup>2</sup> )	FEV <sub>1</sub> Z-score	% predicted FEV <sub>1</sub>	FEV <sub>1</sub> (L)	FEV <sub>1</sub> /height <sup>2</sup> (L/m <sup>2</sup> )	FEV <sub>1</sub> Z-score	% predicted FEV <sub>1</sub>	FEV <sub>1</sub> (L)	FEV <sub>1</sub> /height <sup>2</sup> (L/m <sup>2</sup> )	FEV <sub>1</sub> Z-score	% predicted FEV <sub>1</sub>	FEV <sub>1</sub> (L)	FEV <sub>1</sub> /height <sup>2</sup> (L/m <sup>2</sup> )	FEV <sub>1</sub> Z-score
Sample size	432	893	514	1005	1582	1840	810	1021	1152	892	1149	1043	606	147	1299	703
Mean age in years (SD)	58.7 (8.8)	57.3 (7.7)	55.6 (7.0)	60.7 (9.0)	61.7 (9.0)	63.4 (8.6)	60.3 (8.4)	63.5 (9.2)	64.4 (8.3)	65.7 (7.9)	63.9 (8.4)	64.7 (8.4)	64.0 (7.7)	65.0 (7.8)	65.6 (7.9)	60.6 (7.1)
% Male	49.5	77.6	72.8	51.0	50.8	49.5	53.2	53.7	57.4	44.6	51.1	58.2	59.4	25.2	49.7	52.6
% White	73.1	75.0	80.7	70.4	72.9	75.7	71.5	73.0	78.1	77.6	75.9	77.6	81.0	72.1	76.6	85.2
Mean BMI in kg/m <sup>2</sup>	26.6 (5.3)	27.3 (5.2)	26.9 (5.0)	27.4 (5.5)	28.5 (6.1)	28.6 (6.3)	28.4 (6.2)	28.5 (6.2)	28.1 (6.5)	26.8 (6.4)	28.8 (6.4)	28.3 (6.6)	25.6 (5.8)	24.3 (5.5)	26.6 (6.2)	26.0 (6.0)
Mean BODE index (SD)	0.6 (0.9)	0.9 (1.1)	0.6 (1.0)	0.8 (1.1)	1.6 (1.4)	2.6 (1.6)	1.3 (1.3)	2.1 (1.4)	4.0 (1.4)	4.8 (1.4)	2.7 (1.4)	3.9 (1.4)	5.5 (1.2)	6.1 (1.2)	4.8 (1.4)	5.3 (1.3)
Mean Resting SpO <sub>2</sub> % (SD)	96.8 (2.3)	96.3 (2.3)	96.6 (2.3)	96.4 (2.5)	95.6 (2.8)	95.0 (3.2)	96.0 (2.6)	95.2 (2.9)	94.2 (3.6)	93.5 (4.1)	94.9 (3.2)	94.1 (3.7)	92.8 (4.6)	91.8 (5.4)	93.3 (4.3)	93.1 (4.5)
Health Status, mean (SD)	3.4 (0.9)	3.2 (0.9)	3.3 (0.9)	3.3 (0.9)	2.9 (0.9)	2.8 (0.9)	3.0 (0.9)	2.8 (0.9)	2.7 (0.9)	2.6 (0.9)	2.8 (0.9)	2.7 (0.9)	2.4 (0.9)	2.2 (0.9)	2.5 (0.9)	2.4 (0.9)

