Assessing bronchodilator response by changes in per cent predicted forced expiratory volume in one second

Octavian C Ioachimescu ![ORCID],[1,2] Jose A Ramos,[3] Michael Hoffman,[3] Kevin McCarthy,[3] James K Stoller[3]

[1]Medicine, Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Emory University School of Medicine, Atlanta, Georgia, USA
[2]Sleep Medicine, Atlanta VA Medical Center, Decatur, Georgia, USA
[3]Respiratory Institute, Cleveland Clinic Foundation, Cleveland, Ohio, USA

Correspondence to
Dr Octavian C Ioachimescu, Medicine, Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Emory University School of Medicine, Atlanta, GA 30303-3073, USA; oioac@yahoo.com

ABSTRACT

In pulmonary function testing by spirometry, bronchodilator responsiveness (BDR) evaluates the degree of volume and airflow improvement in response to an inhaled short-acting bronchodilator (BD). The traditional, binary categorization (present vs absent BDR) has multiple pitfalls and limitations. To overcome these limitations, a novel classification that defines five categories (negative, minimal, mild, moderate and marked BDR), and based on % and absolute changes in forced expiratory volume in 1 s (FEV_1), has been recently developed and validated in patients with chronic obstructive pulmonary disease, and against multiple objective and subjective measurements. In this study, working on several large spirometry cohorts from two different institutions (n=31 598 tests), we redefined the novel BDR categories based on delta post-BD–pre-BD FEV_1 % predicted values. Our newly proposed BDR partition is based on several distinct intervals for delta post-BD–pre-BD % predicted FEV_1 using Global Lung Initiative predictive equations. In testing, training and validation cohorts, the model performed well in all BDR categories. In a validation set that included only normal baseline spirometries, the partition model had a higher rate of misclassification, possibly due to unrestricted BD use prior to baseline testing. A partition that uses delta % predicted FEV_1 with the following intervals ≤0%, 0%–2%, 2%–4%, 4%–8% and >8% may be a valid and easy-to-use tool for assessing BDR in spirometry. We confirmed in our cohorts that these thresholds are characterized by low variance and that they are generally gender-independent and race-independent. Future validation in other cohorts and in other populations is needed.

INTRODUCTION

Spirometry is the most commonly used pulmonary function test (PFT), providing objective measurements for diagnosis of lung disease, for global or perioperative risk assessment and for monitoring respiratory health. One PFT modality is represented by the dynamic assessment before and after a bronchodilator (BD), or bronchodilator responsiveness (BDR) testing, which evaluates the degree of volume and airflow improvement in response to an inhaled

Significance of the study

What is already known about this subject?
► Spirometry is the most commonly used pulmonary function test.
► In spirometry, the dynamic assessment before and after a bronchodilator (bronchodilator responsiveness (BDR)) determines the degree of airflow improvement in response to an inhaled bronchodilator such as albuterol.
► Standard, binary BDR categorization (positive or negative) is based on meeting simultaneously an absolute and a % increase from baseline in either forced expiratory volume in 1 s (FEV_1) or in forced vital capacity.
► A novel, non-binary BDR classification defining five distinct categories has been recently developed against several patient-relevant outcomes, and based only on changes in FEV_1 (both absolute and % improvements).

What are the new findings?
► In this study, we correlated the new categories of negative, minimal, mild, moderate or marked BDR with changes in % predicted values of FEV_1.
► The delta % predicted values of FEV_1 is less influenced by anthropometric factors such as height, weight, gender and race than absolute or % changes.

How might these results change the focus of research or clinical practice?
► The cut-offs of the BDR partition based on delta % predicted FEV_1 are gender-independent and race-independent, which allows for an easy-to-use, simplified BDR assessment for all tested subjects.
► If validated in other populations and against other objective and subjective patient-centric outcomes, this new categorization may have a significant impact on the way we diagnose and treat prevalent disorders such as asthma and chronic obstructive pulmonary disease.

short-acting BD such as albuterol. If the aim of the test is to determine whether the spirometric lung function can be improved with therapy in addition to the usual regimen, the subject may continue usual BD medications before the test. If the test is used for diagnosis or to determine whether there is any change in lung function in response to BD, the clinician ordering spirometry should instruct the patient to withhold other BD medications before baseline testing.[1]

The American Thoracic Society (ATS)-European Respiratory Society (ERS) joint guidelines for spirometry define a 'positive' BDR as an absolute 0.2 L and a 12% increase from baseline in either forced expiratory volume in 1 s (FEV_1) or in forced vital capacity (FVC); if neither criterion is met, BDR is classified as 'negative'.[2] From a practical perspective, this categorization has several limitations. For example, those with low FEV_1 or FVC at baseline may not meet the absolute change or delta (Δ) ≥ 0.2 L criterion, while those with preserved lung function (large volumes) at baseline may fail the $\geq 12\%$ rule.[3–5] Over the years, multiple authors[6–8] pointed out that the % change to BD is a continuous variable, and that a single threshold may not separate optimally responders from non-responders.

In order to overcome some of these limitations, Hansen et al[9] recommended recently a novel, non-binary BDR classification, based only on FEV_1, and on absolute or % increases from baseline. The authors differentiated between negative, minimal, mild, moderate and marked responses by using the following thresholds[9] and the most severe impairment criterion[10]: ≤ 0 cL or %, (0, 9] cL or %, (9, 16] cL or %, (16, 26] cL or % and >26 cL or %, respectively (0.01 L=1 cL=10 mL). The study assessed the ability of the novel BDR classes to stratify patient-relevant outcome measures and objective assessments, such as chronic obstructive pulmonary disease (COPD) exacerbation frequency, dyspnea scores, exercise performance, quality of life measurements and radiological airway measurements.[9]

While BDR is generally assessed using absolute and/or % changes from baseline, another possible categorization is by Δ post-BD–pre-BD % predicted values.[11 12] The latter has been recently shown to avoid gender-based and size-based biases in assessing BDR.[11] In order to ascertain if this strategy could provide an easier way to classify BDR, we assess here the relationship between Δ % predicted between pre-BD and post-BD values of FEV_1, FVC and/or FEV_1/FVC ratio, and both standard, binary ATS/ERS and novel BDR classes, in several large PFT cohorts from two different healthcare systems.

METHODS

The study cohorts included all consecutive and acceptable spirometries performed on adult subjects who underwent same-day pre-BD and post-BD measurements at two different institutions and during prespecified periods of time, that is, Cleveland Clinic, in Cleveland, Ohio (n=20,687, 1993–2004 and n=727, 2019–2020) and Atlanta Veteran Affairs Healthcare System in Atlanta, Georgia (AVAHCS, n=4330, 2009–2015 and n=5854, 2015–2020). We organized them as follows: the initial Cleveland Clinic cohort (n=20,687) and the initial AVAHCS cohort (n=4330) were mixed together and constituted the training (random 66%) and the testing (random 33%) sets; the subsequent AVAHCS cohort (n=5854) became the validation set 1, while the most recent Cleveland Clinic cohort (n=727, which included only non-smoking adults with normal FEV_1, FVC and FEV_1/FVC) became the validation set 2.

Spirometry was performed using a Jaeger MasterLab system (Wurzberg, Germany). The ATS/ERS standards and criteria for validity and acceptability[13–15] were used. The post-BD measurements were obtained within 30 min after a standard total dose of 360 μg of inhaled albuterol was administered.

Per the latest ATS/ERS technical statement on spirometry,[1] if the BDR test is done to determine if lung function can be improved above and beyond the existing treatment regimen, the patient may continue taking the usual BD medications before the assessment; if the test is used for diagnosis or to determine whether there is any significant change in lung function in response to BD, then the clinician ordering spirometry should instruct the patient to withhold BD before baseline testing for specific periods of time that are highly dependent on the half-lives of the respective medications.[1] As such, in the Cleveland Clinic 1993–2004 and the AVAHCS 2009–2015 cohorts (together constituting the training and the testing sets), administration of short-acting (albuterol) and long-acting (salmeterol, formoterol) beta-adrenergic BD agents was discouraged within 6 and 24 hours, respectively; short-acting (ipratropium) and long-acting (tiotropium) antimuscarinic agents were recommended to be held before the test for a minimum of 8 and 24 hours, respectively (although neither standardized for all PFT prescribers, nor enforced). No individuals were on ultra long-acting beta-adrenergic (indacaterol, olodaterol, vilanterol) or antimuscarinic (glycopyrrolate, umeclidinium, aclidinium) agents in these cohorts. In the more recent PFT groups, the BD inhalers were withheld for 8–24 hours in the AVAHCS 2015–2020 (validation set 1, based on the specific pharmacokinetics), while other BD were completely unrestricted and patients continued to take them as usual in the Cleveland Clinic 2019–2020 cohort (validation set 2).

The most recent and widely applicable equations for normal lung function, that is, Global Lung Initiative (GLI) splines were used for spirometry evaluation.[16] Normal spirometry was defined as observed values of FEV_1, FVC and FEV_1/FVC between lower and upper limits of normal, as defined by the GLI equations.

Descriptive statistical analysis of study variables was performed. Categorical variables were presented as counts or percentages, and compared by using χ^2 test. Continuous variables were characterized as median and 25th–75th IQR due to non-normality, and compared using Tukey-Kramer honestly significant difference with or without Welch's correction, Wilcoxon, Kruskal-Wallis rank sum or Kolmogorov-Smirnov tests, as appropriate. Exploratory recursive decision trees of up to 10 splits were developed in the training set and subsequently assessed in the testing set, with external validation in the remaining PFT groups, defined a priori (validation sets 1 and 2). The decision trees fitted the response value of novel BDR as categorical variables by Δ % predicted FEV_1, FVC and FEV_1/FVC ratio as continuous variables. After rounding to the next integers, the best models were then selected based on the aims of maximizing entropy and generalized R^2 and the area under operating characteristic curve (AUROC) values, while

Table 1 Functional measurements before and after bronchodilator administration in the PFT study groups

Characteristic	Group	Pre-BD Median (IQR)	Post-BD Median (IQR)	Mean delta change (95% CI) Mean % change (95% CI)
FEV$_1$	A	1.4 (0.9–2.1)	1.5 (2.0–2.2)	0.07 (0.069 to 0.074) 5.2 (–1.0 to 12.0)
	B	2.3 (1.7–2.8)	2.4 (1.8–2.9)	0.12 (0.118 to 0.129) 5.0 (4.7 to 5.2)
	C	2.9 (2.4–3.7)	3.0 (2.4–3.8)	0.1 (0.008 to 0.154) 2.9 (–5.6 to 10.6)
FEV$_1$ % predicted (GLI)	A	44 (30–59)	46 (32–61)	1.9 (1.807 to 1.966) 3.4 (3.2 to 3.6)
	B	71 (56–84)	75 (60–88)	3.8 (3.6 to 3.9) 5.0 (4.7 to 5.2)
	C	95 (87–105)	99 (91–108)	2.5 (1.5 to 3.5) 1.8 (0.4 to 3.3)
FVC	A	2.9 (2.2–3.7)	3.0 (2.3–3.8)	0.08 (0.074 to 0.081) 3.0 (–2.3 to 9.0)
	B	3.4 (2.8–4.1)	3.5 (2.9–4.2)	0.1 (0.095 to 0.109) 2.5 (2.3 to 2.7)
	C	3.9 (3.0–4.8)	3.9 (3.0–4.9)	0.03 (–0.055 to 0.111) 0.1 (–7.1 to 9.0)
FVC % predicted (GLI)	A	72 (57–87)	73 (60–88)	1.8 (1.677 to 1.898) 2.1 (1.9 to 2.3)
	B	82 (71–95)	85 (73–98)	2.4 (2.2 to 2.6) 2.5 (2.3 to 2.7)
	C	98 (90–107)	99 (91–107)	0.6 (–0.4 to 1.6) –0.3 (–1.8 to 1.1)
FEV$_1$/FVC	A	0.53 (0.41–0.61)	0.54 (0.41–0.63)	0.01 (0.009 to 0.010) 1.9 (–2.7 to 6.3)
	B	0.69 (0.58–0.76)	0.71 (0.60–0.78)	0.018 (0.017 to 0.019) 2.4 (2.2 to 2.6)
	C	0.77 (0.72–0.81)	0.81 (0.75–0.84)	1.6 (1.036 to 2.144) 0.1 (–1.9 to 1.9)
FEV$_1$/FVC % predicted (GLI)	A	65 (50–75)	66 (50–76)	0.982 (0.909 to 1.055) 1.0 (0.9 to 1.1)
	B	88 (75–97)	91 (77–99)	2.3 (2.2 to 2.5) 2.4 (2.2 to 2.6)
	C	97 (92–101)	100 (95–104)	0.1 (–0.2 to 0.4) 1.4 (0.6 to 2.3)

A: training and testing combined sets, n=25 017; B: validation set 1 from AVAHCS, n=5854; C: validation set 2 from Cleveland Clinic, n=727.
AVAHCS, Atlanta Veteran Affairs Healthcare System in Atlanta, Georgia; BD, bronchodilator; FEV$_1$, forced expiratory volume in 1 s; FVC, forced vital capacity; GLI, Global Lung Initiative; PFT, pulmonary function test.

minimizing the number of splits (chosen: four to match the number of BDR categories), square root of the mean squared prediction error, mean absolute deviance and misclassification rates. Analyses and graphics were performed using JMP Pro15 (SAS Institute, Cary, North Carolina, USA).

RESULTS

The training and testing sets together included 25 017 consecutive, reproducible and acceptable, dual pre-BD/post-BD spirometry sets from the Cleveland Clinic 1993–2004 and the 2009–2015 AVAHCS cohorts. Tested subjects had a median (IQR) age of 62 (52–70) years. Approximately 35% of the subjects were women. By ethnicity, 79% were white and 20% were black. Median (IQR) body mass index (BMI) was 27 (23–31) kg/m^2.

The validation set 1 (AVAHCS 2015–2020 cohort) had 5854 pre-BD and post-BD tests on subjects 61 (52–67) years of age; 11% were women; 51% were white and 48% black; BMI was 29 (26–33) kg/m^2.

The validation set 2 (Cleveland Clinic 2019–2020 cohort) included 727 adults, 55 (42–68) years of age; 32% were women; 71% white, 17% black and 15% or other races or ethnicities; BMI was 31 (26–36) kg/m^2.

Approximately 21%, 23% and 21% of the training/testing, validation sets 1 and 2 met the standard ATS/ERS 'positive' BDR criteria, respectively. In the training and testing sets (chosen randomly with a preset partition rate of 2:1, hence without significant differences between them), the new BDR categorization included 29%, 24%, 18%, 16% and 13% negative, minimal, mild, moderate or
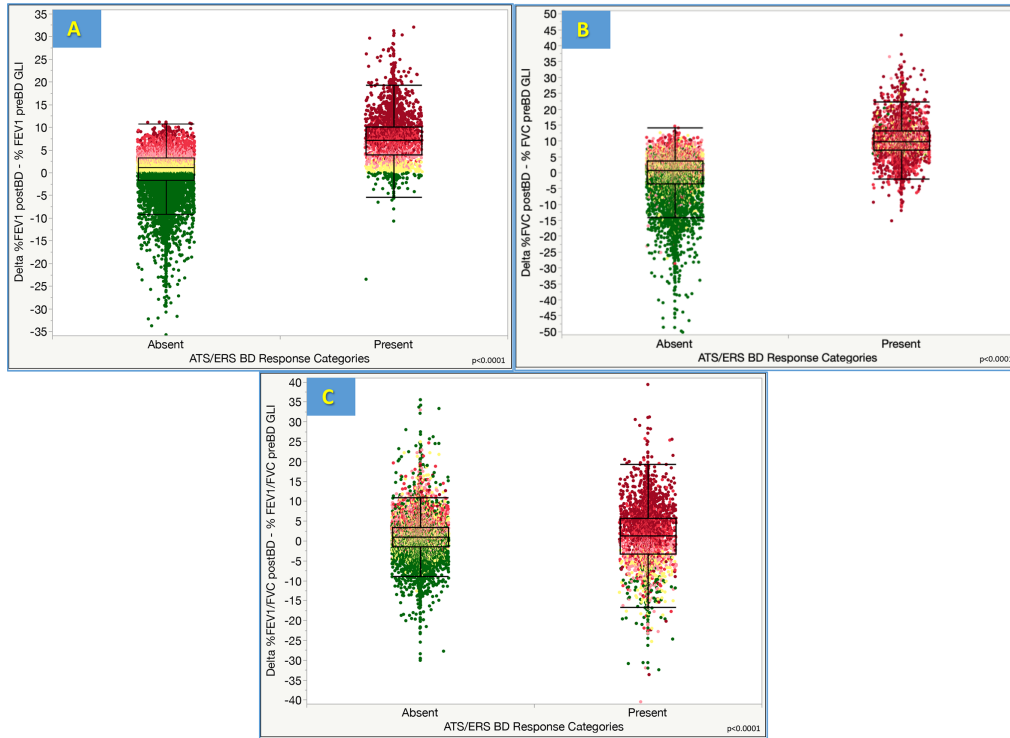
Figure 1 Box-and-whisker plots representing delta post-BD–pre-BD % predicted FEV$_1$ (A), FVC (B) and FEV$_1$/FVC ratio (C) in the standard ATS/ERS BDR categories. Marker color coding was done based on the new BDR categories: green—negative, yellow—minimal, pink—mild, bright red—moderate and dark red—marked. ATS, American Thoracic Society; BD, bronchodilator; BDR, bronchodilator responsiveness; ERS, European Respiratory Society; FEV$_1$, forced expiratory volume in 1 s; FVC, forced vital capacity; GLI, Global Lung Initiative.

marked BDR, respectively. The validation set 1 included 22%, 21%, 17%, 19% and 21% negative, minimal, mild, moderate or marked BDR, respectively; while the validation set 2 had 42%, 9%, 7%, 7% and 35% in the same categories, respectively. table 1 shows the functional parameters studied in the different PFT sets.

Figure 1A–C illustrate the distribution of the differences post-BD–pre-BD for % predicted FEV$_1$, FVC and/or FEV$_1$/FVC ratios, respectively by standard ATS/ERS BDR categories, while figure 2A–C show the same functional parameters by the new BDR categories, all in the testing and training sets together. Online supplemental figure S2
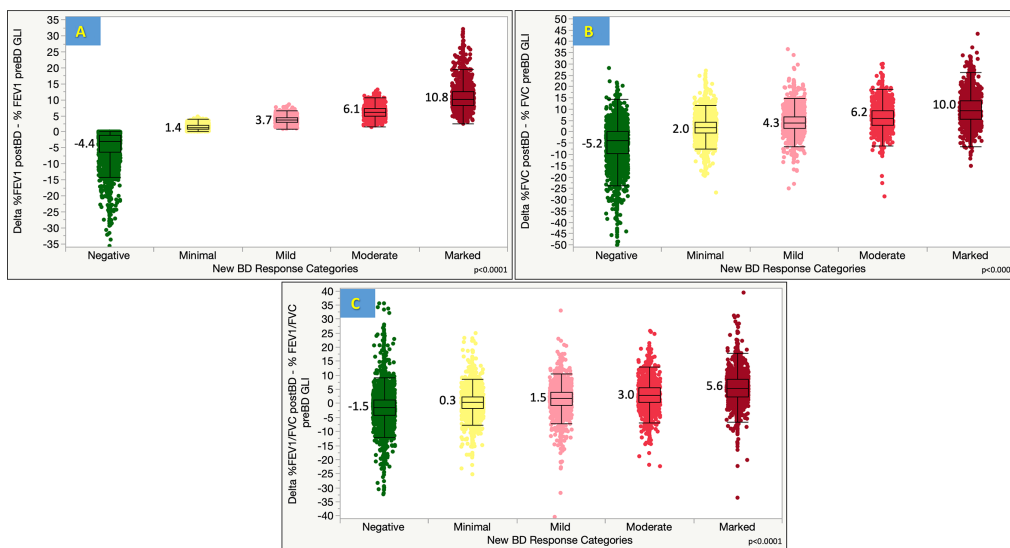

Figure 2 Box-and-whisker plots of delta post-BD–pre-BD % predicted FEV$_1$ (A), FVC (B) and FEV$_1$/FVC ratio (C) against the new BDR categories. Marker color coding was done based on the new BDR categories: green—negative, yellow—minimal, pink—mild, bright red—moderate and dark red—marked. BD, bronchodilator; BDR, bronchodilator responsiveness; FEV$_1$, forced expiratory volume in 1 s; FVC, forced vital capacity; GLI, global lung initiative.
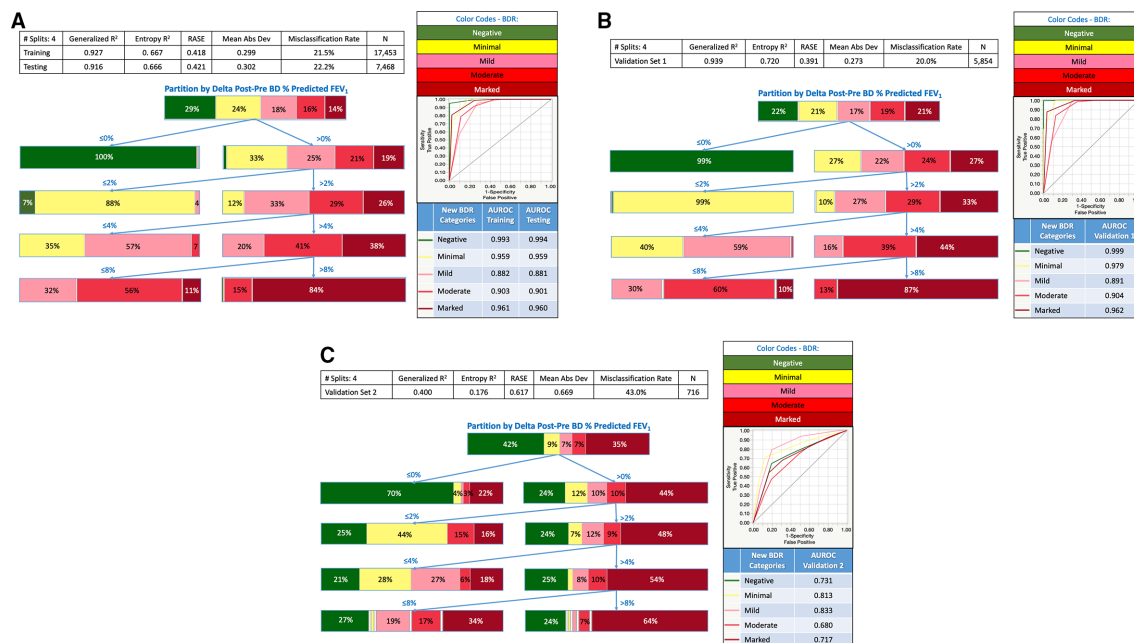
Figure 3 (A) Partition by delta % predicted FEV$_1$ (training set), with random, internal cross-validation in 33% of the cohort (testing set): good partition model (high values for both generalized and entropy R^2). (B) Partition by delta % predicted FEV$_1$ in validation set 1, from AVAHCS (preserved model performance). (C) Partition by delta % predicted FEV$_1$ in validation set 2, from Cleveland Clinic (a decrement in model performance is noted). AUROC, area under receiver operating characteristic curve; AVAHCS, Atlanta Veteran Affairs Healthcare System; BD, bronchodilator; BDR, bronchodilator responsiveness; FEV$_1$, forced expiratory volume in 1 s; Mean Abs Dev, mean absolute deviance; RASE, square root of the mean squared prediction error.

show in the same sets the box-and-whisker plots of mean Δ post-BD–pre-BD % predicted FEV$_1$ (online supplemental figure S1A), FVC (online supplemental figure S1B) and FEV$_1$/FVC ratio (online supplemental figure S1C) against the new BDR categories in the standard ATS/ERS categories of 'present' or 'absent' BDR.

In the data sets studied, the Δ % predicted FEV$_1$ was either statistically similar or clinically insignificant when compared by gender or race. For example, mean Δ % predicted FEV$_1$ was 2.6%–3.8% in men and 1.9%–3.5% in women; by race, it was 1.6%, 2.1%–4.0%, 2.7%–3.5%, 3.3% and 3.6% in north-east Asian, white, black, south-east Asian or in other categories, respectively. When analyzed separately, those self-identified as Hispanic or Latino, had a mean Δ % predicted FEV$_1$ of 3.9%. Furthermore, size measurements such as weight, height and BMI did not

influence in any significant way the variance of the Δ % predicted FEV$_1$ (R^2 <0.01).

We illustrate in figure 3A the proposed partition based on the five intervals for Δ post-BD–pre-BD % GLI-predicted FEV$_1$ and the specific distribution of BDR categories in each interval. The model's generalized R^2 was >0.92, entropy R^2 was high (~0.67), the AUROC was >0.88 in all BDR categories, and the misclassification rates were ~22%. Table 2 shows the definitions of the model's performance metrics in both testing and training sets. Tables 3 and 4 illustrate the confusion matrices for the predicted versus actual BDR categories using the new partition system (perfect correlation is represented by the main diagonal) in the training and testing sets, respectively. Figure 3B,C illustrate the details of the partition and the performance of the new BDR partition in the two validation sets, while tables 5 and 6 show the

Table 2 Partition model's performance using % predicted change in FEV$_1$

Measure	Training	Validation	Definition
Entropy R^2	0.715	0.715	$1 - \text{Loglike(model)}/\text{loglike(0)}$
Generalized R^2	0.934	0.934	$(1 - (L(0)/L(\text{model}))^{(2/n)})/(1 - L(0)^{(2/n)})$
Mean −log p	0.446	0.445	$\sum -\text{Log}(\rho[j])/n$
RASE	0.381	0.382	$\sqrt{\sum (y[j] - \rho[j])^2/n}$
Mean Abs Dev	0.268	0.268	$\sum |y[j] - \rho[j]|/n$
Misclassification rate	0.20	0.21	$\sum(\rho[j] \neq \rho\text{Max})/n$
N	17 543	7, 4, 68	n

FEV$_1$, forced expiratory volume in 1 s; Mean Abs Dev, mean absolute deviance; RASE, square root of the mean squared prediction error.

Table 3 Confusion matrix in the training set

Actual	Training set predicted count				
New BDR category	Negative	Minimal	Mild	Moderate	Marked
Negative	4833	222	0	0	0
Minimal	3	3059	1196	0	0
Mild	5	104	3096	153	3
Moderate	4	1	1191	1133	399
Marked	6	0	30	246	1859

BDR, bronchodilator responsiveness.

Table 5 Confusion matrix in the validation set 1

Actual	Testing set predicted count				
New BDR category	Negative	Minimal	Mild	Moderate	Marked
Negative	1300	4	0	0	0
Minimal	0	830	379	4	0
Mild	0	1	561	456	0
Moderate	0	0	16	916	163
Marked	0	0	0	152	1072

BDR, bronchodilator responsiveness.

confusion matrices for the predicted versus actual BDR in the same validation sets. Overall, the model showed excellent performance in the validation set 1 (generalized R^2 ~0.94, entropy R^2 ~0.72, AUROC >0.89 and misclassification rate of ~20%). Perhaps expectedly, the validation set 2 had a lower performance (generalized R^2 ~0.40, entropy R^2 ~0.18, AUROC >0.68 and misclassification rate of 43%, figure 3C) in the 2019–2020 Cleveland cohort, in which participants were allowed to continue uninterrupted the use of BD prior to the test, likely reducing the overall magnitude of the effect induced by BD administration (together with the normal lung function, ie, large exhaled volumes at baseline). Indeed, in the testing and training sets, the mean Δ post-BD–pre-BD % predicted FEV_1 was −9%, 3.7%, 9.6%, 15.3% and 26.1%, in the validation set 1 it was −4.5%, 1.5%, 3.9%, 6.5% and 12.2%, while in the validation set 2 it was −3.1%, 2.3%, 3.3%, 4.4% and 12.6% in the negative, minimal, mild, moderate and marked BDR categories, respectively (online supplemental figure S2). The SEs of the means for the Δ post-BD–pre-BD % predicted FEV_1 was 0.1%–0.3%, 0.1%–0.2% and 1.1%–1.8% in the testing/ training, validation set 1 and validation set 2, respectively (online supplemental figure S2). When assessed for intrinsic variation or intertest reliability in a subgroup of 17 subjects from the validation set 2 who underwent multiple pre-BD (2–28) and post-BD (2–26) trials on several testing days, the median (IQR) coefficients of variation for Δ post-BD–pre-BD % predicted FEV_1 were very low, that is, 2.9% (1.3–3.3) and 3.2% (1.6–3.2), respectively.

DISCUSSION
We propose here that a partition into five intervals, that is, ≤0, (0–2], (2–4], (4–8] and >8% for delta % predicted FEV_1 is a valid and easy-to-use partition of BDR in spirometry. We correlated these subgroups with the novel BDR categories proposed by Hansen et al,[9] which were developed against various objective and subjective measurements done in patients

with COPD. Further validation in other cohorts and against other objective and subjective assessments is needed, while elucidating the impact of the practice to allow usual inhaler administration prior to BDR testing on this categorization.

Interpretation of BDR in spirometry in patients with airflow limitation or obstruction has been a matter of significant debate for many decades.[17–20] Previously called 'reversibility testing', BDR is a determination of the degree of improvement in flows and volumes after administration of a short-acting inhaled BD such as albuterol. In 1991, an ATS committee recommended using an increase in either FEV_1 or FVC of ≥0.2 L and ≥12% for a significant BDR[15]; this set of criteria was endorsed again in the 2005 ATS/ERS guidelines.[2]

From a practical perspective, the ATS/ERS categorization of 'positive' versus 'negative' BDR categorization[2] has several limitations: it does not always identify clinically significant BDR, it fails to unequivocally partition obstructive lung disorders such as asthma, COPD, asthma-COPD overlap (ACO) and so on, and does not provide therapeutic guidance. For example, which patient should receive a specific medication, from a certain class of BD? Furthermore, those with low FEV_1 or FVC at baseline may not meet the absolute change or delta (Δ) ≥0.2 L criterion, while those with good lung function (high values for FEV_1 or FVC at baseline) may fail the ≥12% rule.[3–5] Hansen et al,[5] analyzing BDR in a sample of 313 tests, found that >70% failed ATS/ERS FEV_1 criteria, while ~40% of those who failed showed statistically significant ΔFEV_1 ≥0.1 L or ~6% improvement. Of those with pre-BD FEV_1 <1 L, more than half had Δ FEV_1 ≥0.1 L or ~6% increase, whereas only 11.4% were 'positive' by ATS/ERS criteria.[4] It has been previously asserted that a 6%–7% change in FEV_1 may represent a significant threshold because it usually corresponds to a mean 0.09–0.10 L increase in FEV_1,[4] which has been suggested to be the minimal clinically important difference for FEV_1.[21] Several authors[6–8] have also pointed out that the % change in response to a BD constitutes a continuous variable and that a single

Table 4 Confusion matrix in the testing set

Actual	Testing set predicted count				
New BDR category	Negative	Minimal	Mild	Moderate	Marked
Negative	2099	118	0	0	0
Minimal	3	1294	534	0	0
Mild	2	37	1266	36	1
Moderate	1	2	491	488	191
Marked	4	0	12	122	767

BDR, bronchodilator responsiveness.

Table 6 Confusion matrix in the validation set 2

Actual	Testing set predicted count				
New BDR category	Negative	Minimal	Mild	Moderate	Marked
Negative	192	29	0	0	78
Minimal	10	44	0	0	8
Mild	3	19	0	0	26
Moderate	9	12	0	0	32
Marked	60	22	0	0	172

BDR, bronchodilator responsiveness.

threshold does not separate optimally responders from non-responders. Considering that the baseline FEV_1 values of individuals tested for BDR vary widely,[4] overcoming healthy population-based CIs[22] for both volumes and % changes may be too restrictive.

In order to improve some of these limitations, a novel BDR grading system based only on FEV_1, and on the highest impairment in volume and % change from baseline was developed: negative ($\leq 0\%$ or ≤ 0 cL), minimal ((0%–9%] or (0–9 cL]), mild ((9%–16%] or (9–16 cL]), moderate ((16%–26%] or (16–26 cL]) and marked (>26% or >26 cL) groups.[9] [10] One centiliter equals 0.01 liter or 10 milliliters. In their investigation on a subgroup of the COPD-Gene study,[23] the authors found negative, minimal, mild, moderate and marked BDR in approximately 21%, 28%, 20%, 18% and 13% of tests, respectively.[9] This BDR distribution closely resembled our BDR categories in the combined Cleveland Clinic-AVAHCS combined cohort, which placed 29%, 24%, 18%, 16% and 13% of the 25 017 tests in the same categories.

While the categorization proposed recently by Hansen *et al* and validated in patients with COPD[9] requires further validation in other populations, especially in its ability to predict daily symptomatic burden, patient-relevant impairments and long-term outcomes of participants with obstructive lung disorders, this new BDR categorization schema may prove to be of major importance in defining ACO and other 'fuzzy' phenotypes of respiratory conditions characterized by airflow limitation. In addition to this classification schema, we propose further investigation and validation of Δ % predicted FEV_1, FVC and FEV_1/FVC ratio between baseline and post-BD state.

As shown above, delta % predicted FEV_1 is a continuous variable that can be divided into five intervals (differentiated by thresholds on an exponential scale), and which accomplishes a BDR partition similar to the one described by Hansen *et al*, based on both absolute and % changes in FEV_1.[9] The variable was confirmed here to be size-independent, gender-independent and race-independent and to separate well tests performed on a routine basis in several PFT laboratories, as well spirometries in two validation cohorts.

In this investigation, we also used a cohort of PFTs performed on non-smoking individuals with normal lung function at baseline (validation set 2), so that we can assess the effect of BD challenge on the Δ % predicted FEV_1 in this population. While the prior use of inhalers was not specifically collected and analyzed in this data set, the observed results do raise the possibility that the specific PFT order to assess BDR with both baseline and post-BD testing unmasks an inherent selection bias (either to rule out an obstructive lung disease or to assess adequacy of treatment in patients with known airflow limitation). The fact that the separation of the BDR categories by the Δ FEV_1 % predicted was less clear, together with the high percentage or marked BDR in the validation set 2 (35%) and the very low 2.4% mean Δ % predicted FEV_1 in the mild and moderate BDR categories, suggests a mix of both the former and the latter scenarios. In addition, it is perhaps not surprising that in subjects with normal lung function at baseline (as in the validation set 2), the % change was not as large as in the case of the tests performed in the PFT laboratory based on routine clinical

indications and specific orders for cases of confirmed or strongly suspected obstructive lung disease.

The current study's strengths are represented by the very large number of PFTs in the various cohorts used for analyses and from two different healthcare systems, the investigative design that included a priori defined testing, training and two validation sets, the fact that the new, 5-group BDR classification schema has been developed against other functional (both objective and subjective) measurements, and the robust results during validation phase, which showed great reproducibility and very small cohort effects. Several of the features of this investigation could be construed as weaknesses: the lack of outcomes data for the subjects tested in these groups, potential distortions induced by some cohort effects (eg, normal, healthy individuals with preserved lung function), the relatively narrow intervals of partition which may not allow fine tuning of the classifications (yet the new definition includes five distinct categories), and the expected finding that unrestricted use of inhalers prior to BDR testing may limit our ability to split optimally the tested subjects into different nosological categories.

CONCLUSIONS

When pre-BD and post-BD spirometry testing is performed, a partition based on ≤ 0, 0–2, 2–4, 4–8 and >8% intervals for delta % predicted FEV_1 is a valid and easy-to-use assessment of BDR. We developed these intervals based on their ability to partition BDR along the same category lines as proposed recently by Hansen *et al*,[9] and found that they are generally size-independent, gender-independent and race-independent. Further validation in other populations and against other objective and subjective assessments is needed, while investigating the extent to which the practice of allowing BD administration prior to the standardized baseline testing influences this categorization and interpretation of the results for diagnostic and therapeutic purposes.

ORCID iD
Octavian C Ioachimescu http://orcid.org/0000-0001-9047-6894

REFERENCES

1 Graham BL, Steenbruggen I, Miller MR, *et al*. Standardization of spirometry 2019 update. An official American thoracic Society and European respiratory Society technical statement. *Am J Respir Crit Care Med* 2019;200:e70–88.

2 Pellegrino R, Viegi G, Brusasco V, *et al*. Interpretative strategies for lung function tests. *Eur Respir J* 2005;26:948–68.

3 Hansen JE, Casaburi R, Goldberg AS. A statistical approach for assessment of bronchodilator responsiveness in pulmonary function testing. *Chest* 1993;104:1119–26.

4 Hansen JE, Porszasz J. Rebuttal from Drs Hansen and Porszasz. *Chest* 2014;146:542–4.

5 Hansen JE, Sun XG, Adame D, *et al*. Argument for changing criteria for bronchodilator responsiveness. *Respir Med* 2008;102:1777–83.

6 Pellegrino R, Brusasco V. Rebuttal from Drs Pellegrino and Brusasco. *Chest* 2014;146:541–2.

7 Calverley PMA, Burge PS, Spencer S, *et al*. Bronchodilator reversibility testing in chronic obstructive pulmonary disease. *Thorax* 2003;58:659–64.

8 Hay JG, Stone P, Carter J, *et al*. Bronchodilator reversibility, exercise performance and breathlessness in stable chronic obstructive pulmonary disease. *Eur Respir J* 1992;5:659–64.

9 Hansen JE, Dilektasli AG, Porszasz J, *et al*. A new bronchodilator response grading strategy identifies distinct patient populations. *Ann Am Thorac Soc* 2019;16:1504–17.

10 Ioachimescu OC. On a new approach to assess bronchodilator responsiveness. *Ann Am Thorac Soc* 2020. doi:10.1513/AnnalsATS.202008-966RL. [Epub ahead of print: 21 Oct 2020].

11 Ward H, Cooper BG, Miller MR. Improved criterion for assessing lung function reversibility. *Chest* 2015;148:877–86.

12 Brand PL, Quanjer PH, Postma DS, *et al*. Interpretation of bronchodilator response in patients with obstructive airways disease. The Dutch chronic non-specific lung disease (CNSLD) Study Group. *Thorax* 1992;47:429–36.

13 Standardization of spirometry, 1994 update. American thoracic Society. *Am J Respir Crit Care Med* 1995;152:1107–36.

14 Miller MR, Hankinson J, Brusasco V, *et al*. Standardisation of spirometry. *Eur Respir J* 2005;26:319–38.

15 Lung function testing: selection of reference values and interpretative strategies. American thoracic Society. *Am Rev Respir Dis* 1991;144:1202–18.

16 Quanjer PH, Stanojevic S, Cole TJ, *et al*. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012;40:1324–43.

17 Freedman BJ, Meisner P, Hill GB. A comparison of the actions of different bronchodilators in asthma. *Thorax* 1968;23:590–7.

18 Criteria for the assessment of reversibility in airways obstruction. Report of the Committee on emphysema American College of chest physicians. *Chest* 1974;65:552–3.

19 Reis AL. Response to bronchodilators. In: Clausen JL, Abramon JF, eds. *Pulmonary function testing guidelines and controversies: equipment, methods, and normal values*. New York: Academic Press, 1982: 215–21.

20 Eliasson O, Degraff AC. The use of criteria for reversibility and obstruction to define patient groups for bronchodilator trials. Influence of clinical diagnosis, spirometric, and anthropometric variables. *Am Rev Respir Dis* 1985;132:858–64.

21 Donohue JF. Minimal clinically important differences in COPD lung function. *COPD* 2005;2:111–24.

22 Tan WC, Vollmer WM, Lamprecht B, *et al*. Worldwide patterns of bronchodilator responsiveness: results from the burden of obstructive lung disease study. *Thorax* 2012;67:718–26.

23 Regan EA, Hokanson JE, Murphy JR, *et al*. Genetic epidemiology of COPD (COPDGene) study design. *COPD* 2010;7:32–43.