



Grading the severity of airways obstruction: new wine in new bottles

Philip H. Quanjer^{1,2}, Jeffrey J. Pretto^{3,4}, Danny J. Brazzale^{5,6} and Piotr W. Boros⁷

Affiliations:

¹Dept of Pulmonary Diseases, Erasmus Medical Centre, Erasmus University, Rotterdam, and

²Dept of Paediatrics, Erasmus Medical Centre, Erasmus University, Rotterdam, The Netherlands.

³Dept of Respiratory and Sleep Medicine, John Hunter Hospital, Newcastle,

⁴School of Medicine and Public Health, University of Newcastle, Newcastle,

⁵Dept of Respiratory and Sleep Medicine, Austin Hospital, Heidelberg, and

⁶Institute for Breathing and Sleep, Heidelberg, Australia.

⁷Lung Function Lab, National TB and Lung Diseases Research Institute, Warsaw, Poland.

Correspondence:

P.H. Quanjer, Kervel 19, 7443 GT Nijverdal, The Netherlands.

E-mail: pquanjer@ziggo.nl

ABSTRACT The objective of this study was to redesign the current grading of obstructive lung disease so that it is clinically relevant and free of biases related to age, height, sex and ethnic group.

Spirometric records from 17 880 subjects (50.4% female) from hospitals in Australia and Poland, and 21 191 records (53.0% female) from two epidemiological studies (age range 18–95 years) were analysed. We adopted the American Thoracic Society(ATS)/European Respiratory Society (ERS) criteria for airways obstruction based on an forced expiratory volume in 1 s (FEV₁)/(forced) vital capacity ((F)VC) ratio below the fifth percentile and graded the severity of pulmonary function impairment using z-scores for FEV₁, which signify how many standard deviations a result is from the mean predicted value.

Using the lower limit of normal for FEV₁/(F)VC and z-scores for FEV₁ of -2, -2.5, -3 and -4 to delineate severity grades of airflow limitation leads to close agreement with ATS/ERS severity classifications and removes age, sex and height related bias.

The new classification system is simple, easily memorised and clinically valid. It retains previously established associations with clinical outcomes and avoids biases due to the use of per cent predicted FEV₁. Combined with the Global Lung Function prediction equations it provides a worldwide diagnostic standard, free of bias due to age, height, sex and ethnic group.



@ERSpublications

Using FEV₁ z-scores to classify severity of airways obstruction is clinically valid and overcomes bias inherent in % pred <http://ow.ly/pRyGI>

Received: May 21 2013 | Accepted after revision: July 16 2013 | First published online: Aug 29 2013

Conflict of interest: None declared.

Copyright ©ERS 2014

Introduction

It is accepted that a diagnosis of airflow limitation should be based on an abnormally low ratio of the forced expiratory volume in 1 s (FEV₁) and vital capacity (expressed as the larger of inspiratory or expiratory vital capacity (VC), or forced vital capacity (FVC)). Surprisingly, there is no universal agreement on what constitutes a low FEV₁/(F)VC ratio. The American Thoracic Society (ATS)/European Respiratory Society (ERS) recommend the fifth percentile of the distribution in a population of healthy lifelong nonsmokers as the lower limit of normal (LLN) [1–3], whereas the Global Initiative for Chronic Obstructive Lung Disease (GOLD) [4] recommend a post-bronchodilator ratio of <0.70 as indicative of persistent airways obstruction. The latter lower limit was arbitrarily selected because of its simplicity, and because it constituted a limit that would not vary with the prediction equation used. Many investigators have documented that the GOLD criterion for defining airway obstruction can result in marked overestimation of the prevalence of airway obstruction in subjects aged ≥45 years, which may lead to unnecessary medical expenditure and to underestimation in younger adults [5]. This is because using a fixed ratio of 0.70 ignores both the natural decline of the FEV₁/(F)VC ratio with age and the sex differences observed in a normal population. The limitations of this approach and the resultant misdiagnoses have been acknowledged by the GOLD group [4]. A standardised approach to interpretation of spirometry utilising the LLN for FEV₁/(F)VC provides a statistically and scientifically more valid approach to defining abnormality in the FEV₁/(F)VC ratio, with consequent reductions in misdiagnoses of pathological airways obstruction and earlier detection of mild disease.

Once a diagnosis of airways obstruction has been made, spirometry values are commonly used to categorise pulmonary function impairment and the ATS/ERS recommended approach is to use the FEV₁ as a percentage of the predicted value (FEV₁ % pred) as the basis for this classification [3]. Current ATS/ERS recommendations define a FEV₁ % pred of >70% as mild impairment, 60–69% as moderate impairment, 50–59% as moderately severe impairment, 35–49% as severe impairment, and <35% as very severe impairment [3]. This approach uses an arbitrary number of severity categories, using arbitrarily decided cut-off values. However, the use of per cent predicted in this way leads to a pronounced age-related bias. For example, the LLN for FEV₁ in elderly healthy subjects and in preschool children may be as low as 65% pred and 72% pred, respectively [6, 7]. The result is that the severity of impairment category corresponds to differing degrees of abnormality for different age groups. A more valid approach would be to take into consideration the underlying distribution of normal lung function data when classifying test results into severity category. The z-score (signifying the number of standard deviations a result is from the mean predicted value) provides a metric for achieving this. A similar approach has previously been shown to provide effective risk stratification for death from respiratory and vascular causes [8].

The objective of the present study is to evaluate an alternative to the ATS/ERS recommendations for grading respiratory impairment to overcome biases related to age, height and sex.

Materials and methods

Materials

We used adult data from two population-based studies: the Health Survey for England (HSE) comprising 9961 subjects (53.8% female, age range 18–93 years) [9] and the National Health and Nutrition Examination Survey (NHANES III) comprising 11 230 subjects (52.2% female, age range 18–90 years) [10]. All data from these epidemiological studies were included in the analysis, and not just those data used for the derivation of normal predicted values. We also utilised data from three clinical populations: 2776 records from the Austin Hospital in Victoria, Australia (49.8% female, age range 18–95 years); 4258 from the John Hunter Hospital in New South Wales, Australia (53.8% female, age range 18–95 years); and 10 846 from the National Research Institute of TB and Lung Diseases (TLDRI) in Warsaw, Poland (47.8% female, age range 18–92 years). The clinical data comprised consecutively collected test results from patients referred for lung function assessment for the typical wide range of clinical purposes seen in large hospital-based clinical laboratories. Tests were conducted between August 2008 and June 2012 (John Hunter Hospital), January 2001 and May 2012 (Austin Hospital), and April 2009 and June 2012 (TLDRI). All spirometry was performed in accordance with internationally agreed standards applicable at the time of data collection [11, 12], and only baseline or pre-bronchodilator status data was included in the analysis.

All analyses were limited to those aged ≥18 years and to those of European descent, because there were few non-European data in the clinical dataset. This study is a retrospective analysis of de-identified data, obviating the need for approval from local Ethics Committees.

Methods

Predicted values and z-scores were derived for each subject in each dataset using prediction equations from the Global Lung Function Initiative (GLI-2012) [7], using specially developed software (GLI-2012 Data

Conversion software; www.lungfunction.org/files/InstallGLI2012_DataConversion.EXE). The z-score indicates the number of standard deviations that a measurement differs from the mean predicted value. Airway obstruction was diagnosed using the ATS/ERS definition [3] of $FEV_1/(F)VC < LLN$, *i.e.* the z-score was < -1.645 [1–3, 6, 7]. Categorisation of the severity of airway obstruction was also made using the five category scale recommended by ATS/ERS [3].

Data analysis was performed using the statistical software R (Version 3.0.1; The R Project for Statistical Computing; www.r-project.org). As the relationship between the z-score for FEV₁ and FEV₁ % pred is symmetrically heteroscedastic (fig. 1), the mean z-score was derived at ratios of 0.8, 0.7, 0.6, 0.5, 0.4, 0.35 and 0.30, and subsequently described by linear regression. Differences between the ATS/ERS recommended scale and new grading systems were tested using two-sided t-tests and by multinomial logistic regression using sex, age and height categories as covariates.

Results

Table 1 shows the age distribution of the subjects in the five study groups. As expected, the prevalence rate of airways obstruction is much higher in clinical patients than in the population-based samples (fig. 2).

There is a strong linear relationship between FEV₁ % pred and FEV₁ expressed as a z-score (fig. 1) across the FEV₁ 30–80% pred range ($z\text{-score } FEV_1 = 5.589 \times FEV_1 \% \text{ pred} - 5.894$, adjusted $R^2 = 0.9997$). This allows reliable reconstruction of ATS/ERS severity classification cut-off values by simple replacement of FEV₁ % pred with z-scores. After accounting for differences in age there were clinically trivial, albeit statistically significant, differences in intercept: -0.0048 in centre 2 (John Hunter Hospital) and 0.0024 in centre 3 (TLDRI).

Table 2 shows the z-score equivalencies to the ATS/ERS schema. The z-scores have been rounded for simplification since it is cumbersome to use z-scores to precise values and, given that the category boundaries were arbitrarily defined originally [3], this rounding will have little overall effect. Z-scores of -2 , -2.5 , -3 and -4 can be used as cut-off values that faithfully reproduce the ATS/ERS severity categories.

FEV₁ % pred correlated negatively ($R = 0.29$; $p < 0.001$) with age and sex (higher FEV₁ % pred in females). The relationship between FEV₁ % pred and z-score for FEV₁ was significantly affected by age (fig. 1) and sex ($R = 0.988$) (data not shown). At an FEV₁ of 70% pred the z-score for FEV₁ varied between -1.53 and -2.59 ; at a z-score for FEV₁ of -2 FEV₁ % pred varied between 68% pred and 81% pred (fig. 1). In patients with an FEV₁/FVC ratio $< LLN$ the average ATS/ERS grade of 2.79 differed significantly from the 2.69 average grade with the new system (two-sided t-test, $p = 0.0006$). Multinomial logistic regression revealed that allocation to more severe grades of airways obstruction using the ATS/ERS recommendations compared with the new allocation system was significantly related to age, sex and height.

There were 2097 (11.6%) clinical patients classified as having severe or very severe airflow limitation using the GOLD definition of $FEV_1/FVC < 0.70$ and $FEV_1 < 50\%$ pred [4]. The proposed new system classified a similar number as having severe or very severe airflow limitation ($n = 2079$, 11.5%); however, in 624 (29.8%) patients the GOLD and proposed criteria lead to a different classification. The GOLD classification system underestimated severity of airways obstruction in 303 of these patients and overestimated severity in 321. Figure 3 shows the prevalence of severe and very severe airway obstruction using these two classification systems as related to age.

FIGURE 1 Relationship between the z-score for forced expiratory volume in 1 s (FEV₁) and FEV₁ as percentage of the predicted value, with subjects stratified by age group; values above zero z-score and 100% pred fan out in a symmetrical way.

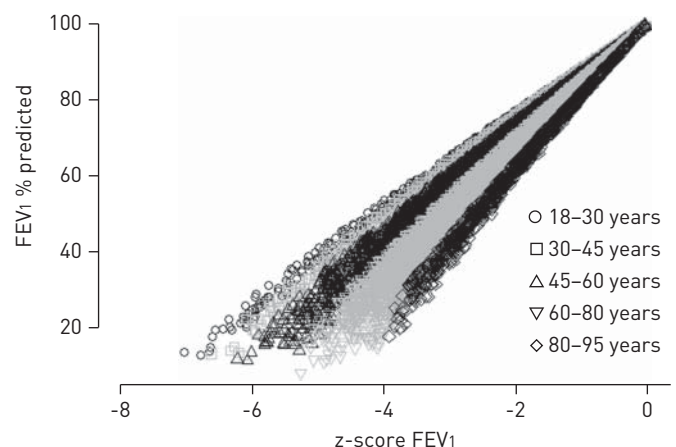


TABLE 1 Number of subjects grouped by age in the five populations included in this study

Sex	Population	Age years								
		18–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–95
Females	AH	8	46	91	127	211	334	340	228	8
	JHH	29	90	119	187	320	517	486	217	2
	TLDRI	34	290	606	600	1412	1485	1029	205	2
	NHANES III	202	1080	925	792	590	779	579	404	19
	HSE	98	586	925	844	904	662	462	110	7
Males	AH	12	55	108	189	281	351	265	118	4
	JHH	26	122	150	253	437	622	479	195	7
	TLDRI	33	292	466	600	1360	1354	893	180	5
	NHANES III	236	1144	1117	844	668	745	657	409	40
	HSE	115	707	1068	1025	1049	719	506	165	9

AH: Austin Hospital, Victoria, Australia; JHH: John Hunter Hospital, New South Wales, Australia; TLDRI: National Research Institute of TB and Lung Diseases, Warsaw, Poland; NHANES III; National Health and Nutrition Examination Survey III; HSE: Health Survey for England 1995–1996.

Table 3 indicates the small differences in severity classification between the ATS/ERS grading system and the new z-score based grading system. The overall prevalence of airways obstruction is identical with the old and new systems since they both utilise FEV₁/FVC <LLN to define abnormality.

Discussion

The primary objective of this study was to design a statistically robust system for detecting airways obstruction and for grading the extent of impairment. Our analysis reveals that the ATS/ERS recommended system of using FEV₁ % pred to grade respiratory impairment is prone to biases caused by age, height and sex differences in underlying distribution of lung function. This can be circumvented by the use of z-scores, since they fully account for these biases, and also for bias that may be introduced by ethnicity factors. The bias due to sex and height, although statistically highly significant, is quite small because the coefficient of variation in males and females is virtually the same [7]. To illustrate this point, a 30-year-old 170 cm subject with a z-score for FEV₁ of -2 has a FEV₁ % pred of 75.9% pred and 75.4% pred for a male or female, respectively; at age 75 years the corresponding values are 65.1% pred and 64.3% pred, respectively.

The age bias has created recent discussion and controversy about the criteria used to define airways obstruction [5, 13–21]. The ATS and ERS advocate use of the fifth percentile of FEV₁/FVC from a healthy population as the LLN [3], whereas the GOLD group adopted a fixed 0.70 cut-off value for the FEV₁/FVC ratio because of its simplicity and independence of different prediction equations [4]. The GOLD criterion for airway obstruction, by ignoring sex differences and the natural decline of the FEV₁/FVC ratio with age to well below 0.70, leads to a >50% overestimation of the prevalence rate of airway obstruction in subjects aged ≥45 years and to significant underestimates in younger adults [5]. These findings are corroborated in

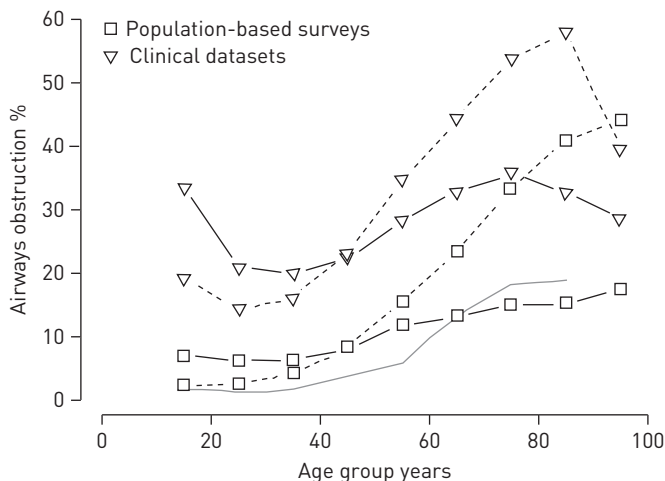


FIGURE 2 Percentage of subjects with airways obstruction for the clinical datasets and population-based surveys using the American Thoracic Society/European Respiratory Society definition of forced expiratory volume in 1 s (FEV₁)/forced vital capacity (FVC) less than the lower limit of normal (solid lines) and using the Global Initiative for Chronic Obstructive Lung Disease definition of FEV₁/FVC <0.70 (dotted lines). The grey line depicts the percentage of healthy nonsmokers with FEV₁/FVC <0.70 derived from the population-based surveys.

TABLE 2 Proposed new grading system for categorisation of airways obstruction

Obstruction	Grade	ATS/ERS 2005	Proposed
Mild	1	>70% pred	$z\text{-score} \geq -2$
Moderate	2	60–69% pred	$-2.5 \leq z\text{-score} < -2$
Moderately severe	3	50–59% pred	$-3 \leq z\text{-score} < -2.5$
Severe	4	35–49% pred	$-4 \leq z\text{-score} < -3$
Very severe	5	<35% pred	$z\text{-score} < -4$
Mean grade		2.79	2.69

The conventional American Thoracic Society (ATS)/European Respiratory Society (ERS) recommended system [3] is based upon defining airways obstruction using forced expiratory volume in 1 s (FEV₁)/forced vital capacity (FVC) <lower limit of normal (LLN) and FEV₁ % predicted to classify severity. The proposed system is based upon defining airways obstruction using FEV₁/FVC <LLN and using z-scores for FEV₁ to classify severity.

the present study (fig. 2) and have been implicitly acknowledged by the GOLD group [4]. On that account we adopted the LLN for FEV₁/(F)VC as the gold standard for diagnosing airways obstruction. It is important to note that this study is not about diagnosing clinical syndromes, such as asthma or chronic obstructive pulmonary disease (COPD) (which rely on clinical assessment in addition to spirometric measurement), but to provide a statistically valid system for interpretation of spirometry.

There is obviously some overlap in FEV₁/FVC ratio between those with and those without respiratory disease, and spirometry results cannot be used to categorically separate ill from healthy individuals. How test results are interpreted is clearly important and the choice between ATS/ERS and GOLD guidelines for this requires some justification. One key difference between the two guidelines is that the GOLD criteria identify a higher incidence of airflow obstruction in elderly subjects. One would expect that subjects in this “marginal zone” (with FEV₁/FVC <0.70 but >LLN) would develop clinical symptoms and signs of disease, and follow-up studies have shed light on this clinically relevant issue. In longitudinal studies of subjects in this marginal zone no association has been found with increased risk of all-cause mortality [13–15, 22] (except in symptomatic smokers [16]), development of respiratory symptoms [16], accelerated decline in FEV₁ [13, 16–18], or with respiratory care utilisation or poorer quality of life scores compared with a reference group [16]. In a review of the literature, MOHAMED HOESEIN *et al.* [23] could only find one publication by MANNINO *et al.* [24] allegedly demonstrating that subjects with an FEV₁/(F)VC ratio <0.70 but >LLN had an increased risk of premature death and hospitalisation for COPD, and on that basis accepted that GOLD grade I does represent respiratory disease. However, MANNINO *et al.* [24] misrepresented their own findings: the adjusted hazard ratio for death of 1.1 was not elevated given the confidence interval for this value was 0.96–1.3. Also, the authors conceded that the “measure of COPD-related hospitalisations was too inclusive” [25]. Thus, there is no evidence that this marginal zone represented by GOLD stage I indicates respiratory disease. Conversely, individuals with FEV₁/(F)VC <LLN, but not nonsmokers with FEV₁/(F)VC <0.70 but >LLN, have an increased risk of all-cause mortality [19, 20], development of respiratory symptoms [19, 20], accelerated decline in FEV₁ [16] and hospitalisation

FIGURE 3 Percentage of subjects with severe or very severe airways obstruction for the clinical dataset using the proposed classification system (forced expiratory volume in 1 s (FEV₁)/forced vital capacity (FVC) < lower limit of normal and z-score for FEV₁ < -3) (solid line) or the Global Initiative for Chronic Obstructive Lung Disease definition (FEV₁/FVC <0.70 and FEV₁ <50% pred) (dotted line).

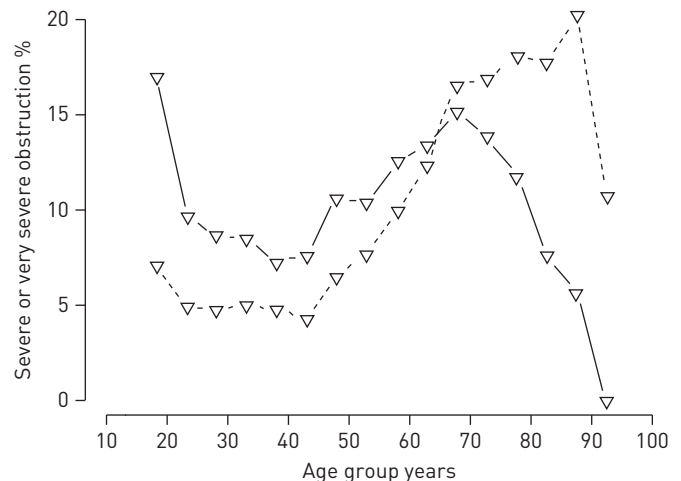


TABLE 3 Subjects falling into the respiratory impairment severity categories as defined using the current American Thoracic Society (ATS)/European Respiratory Society (ERS) guidelines and the proposed new grading system

	ATS/ERS classification					Total
	No obstruction	Mild	Moderate	Moderately severe	Severe	
Clinical data						
New classification						
No obstruction	70.6	1.7	1.0	0	0	73.3
Mild	0	5.9	0.1	0	0	5.9
Moderate	0	0.7	2.6	1.0	0	4.4
Moderately severe	0	0	1.2	2.3	1.3	4.8
Severe	0	0	0.1	1.7	4.3	7.5
Very severe	0	0	0	0	1.1	4.1
Total	70.6	8.3	5.0	5.0	6.6	100.0
Epidemiological data						
New classification						
No obstruction	90.3	0.7	0.3	0	0	91.3
Mild	0	4.9	0	0	0	4.9
Moderate	0	0.4	0.6	0.2	0	1.2
Moderately severe	0	0	0.4	0.4	0.2	1.0
Severe	0	0	0	0.3	0.6	1.2
Very severe	0	0	0	0	0.1	0.4
Total	90.3	6.0	1.3	0.9	0.9	100.0

Data are presented as %. Subjects have been separated into clinical and population-based datasets.

for COPD [21]. Therefore, we believe that there is overwhelming evidence that the LLN, defined as the fifth percentile of the distribution in a reference population, should be used to diagnose pathological airflow limitation.

Another issue of concern is the use of FEV₁ % pred for categorising disease severity. There have been numerous publications, summarised in the GLI-2012 report [7], that show the underlying scatter in FEV₁ is not proportional to the predicted value, and that it varies with age. There is, therefore, a widening gap between 80% pred and the LLN for FEV₁, particularly in adults aged ≥45 years (fig. 4), leading to an age-related overestimate of the severity of airflow limitation. The proposed modifications to the ATS/ERS criteria (table 2) remove the bias in age, sex and height; they led to identification of 4% of healthy females and males from the NHANES III and HSE studies aged ≥18 years as having mild airways obstruction.

Using the absolute value of FEV₁/FVC to define disease severity is problematic, in that more severe airflow obstruction can lead to a reduction in FVC with a paradoxical increase in FEV₁/FVC and poorer prognosis [26].

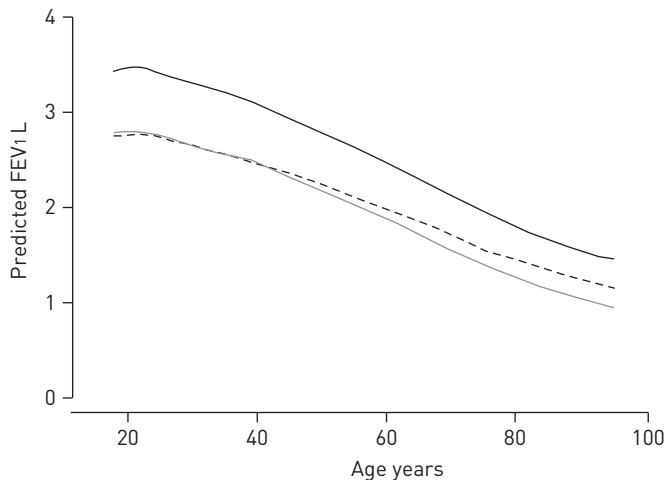


FIGURE 4 Relationship between age and the predicted value for forced expiratory volume in 1 s (FEV₁) (solid black line), its lower limit of normal (solid grey line) and FEV₁ 80% pred (dashed black line) in females aged 18–95 years.

For this reason severity scaling using FEV₁ % pred was recommended [3]. FEV₁/FVC ratios as low as 0.31, 0.36, 0.29, 0.21 and 0.15 occurred in ATS/ERS criteria and were defined as mild, moderate, moderately severe, severe and very severe airway obstruction, respectively. Hence we have focused on FEV₁ z-scores.

Correct classification of obstructive spirometry values in the moderate to severe range is of importance in both clinical and research settings. The most recent GOLD guidelines (2011–2013) [4] introduced categorisation for risk of exacerbation, with one condition leading to patients being categorised as “high risk” if FEV₁ was <50% pred. We identified 2097 (11.6%) patients meeting this spirometric criterion. Nearly the same number of patients were identified with severe or very severe airway obstruction according to the new proposed criteria (FEV₁/FVC < LLN and FEV₁ z-score < -3). Although these results appear to be similar, there is actually a 29.8% disagreement in classification. In the younger age groups the GOLD criterion underestimates the real severity of the obstruction in half of these patients and overestimates it in half of the elderly (fig. 3). This may lead to inappropriate management and/or treatment in both situations and introduces a considerable bias in the selection of patients for intervention studies and other research.

Utilising the underlying spread of spirometry data to stratify risk is not new. In 1983, Peto *et al.* [8] described a 10-fold increase in risk of death from respiratory disease over the 20 subsequent years in those with a FEV₁ more than two standard deviations below average. In developing these new classification criteria we have used z-scores as an essentially identical approach to help stratify severity of disease state. The use of z-scores is already standard procedure when evaluating bone mineral density and growth curves [27, 28], and is recommended for the interpretation of lung function test results [6, 7]. However, recent analyses reveal that neither FEV₁ % pred nor z-scores provide the best indices of choice for studying mortality. Expressing lung function as multiples of the minimum FEV₁ considered to be compatible with life (estimated at 400 mL and 500 mL in female and male adults, respectively) may provide better prediction of mortality [29]. In a 175 cm tall male aged 30 years this minimum FEV₁ of 500 mL represents an FEV₁ of 11.5% pred, z-score -6.51, and at age 75 years an FEV₁ of 17.0% pred, z-score -4.17; illustrating that both the per cent predicted and z-score approaches do not account for this age bias. A z-score of -4.17 in the elderly subject is equivalent to -6.51 in the younger subject from a mortality perspective. Hence, disease severity should be categorised using our suggested cut-off values but we should not be tempted to use absolute z-scores to further stratify disease severity.

Spirometric indices of pulmonary function correlate poorly with respiratory symptoms, dyspnoea, exacerbations, hospitalisation, exercise limitation and quality of life. Therefore, any system for grading severity of airflow limitation based solely on spirometric test results will inherently fail to satisfactorily capture the complete clinical picture. Rather than being treated solely as a disease of the lungs, COPD is now considered by some as a multifactorial systemic disease [30]. In view of this the GOLD committee introduced a new classification system that combines spirometry with respiratory symptoms in an attempt to quantify future risk; however, this system had not been previously validated. LEIVSETH *et al.* [22] concluded that GOLD grade 1 was not associated with increased mortality, and that spirometric GOLD grades 2 and higher predicted mortality better than the new GOLD ABCD groups amongst people with COPD from a Norwegian general population. Also, survival is better in the more severe COPD group C (low lung function but less dyspnoea) than in the less severe group B (much better lung function but more dyspnoea) [31], and the addition of dyspnoea and exacerbations to the severity classification did not add prognostic value on long-term COPD outcomes [32]. It therefore seems that more research is required to clinically phenotype the heterogeneous syndrome of COPD and identify biomarkers of disease, with a view to bringing treatment tailored to the individual patient within reach [33], and improving the classification of respiratory impairment and improving risk assessment.

Adopting the proposed adjustments to the ATS/ERS recommendations avoids contaminating research groups with healthy subjects and avoids unnecessary medical interventions. In addition, this proposal would enable younger patients with milder disease, who would not have been identified using the GOLD criteria, to be included in future research initiatives. The all-age prediction equations for different ethnic groups issued by the Global Lung Function Initiative [7], and endorsed by six large international respiratory societies, have for the first time provided a means for categorising the LLN and also the severity of pulmonary impairment free of bias due to age, height, sex or ethnic group. Thus, a diagnosis of airways obstruction and categorisation of its severity no longer needs to be biased and dependent on where the measurements were performed.

Conclusion

The redesigned system for diagnosing airways obstruction and for categorising pulmonary function impairment is based on evidence which underpins its clinical validity and retains previously documented associations between severity of respiratory impairment and outcome variables, such as all-cause death.

Application with the GLI-2012 prediction equations [7], coupled with standardisation in testing methodologies [11], provides the opportunity for a worldwide standard for performance and interpretation of spirometry, which is free of major sources of bias.

Acknowledgements

The authors gratefully acknowledge the constructive comments from the reviewers and Associate Editor M.R. Miller which helped to substantially improve the manuscript.

References

- 1 American Thoracic Society. Lung function testing: selection of reference values and interpretative strategies. *Am Rev Respir Dis* 1991; 144: 1202–1218.
- 2 Quanjer PH, Tammeling GJ, Cotes JE, et al. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J* 1993; 6: Suppl. 16, 5–40.
- 3 Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948–968.
- 4 Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management and Prevention of COPD. GOLD, 2013. Available from: www.goldcopd.org/guidelines-global-strategy-for-diagnosis-management.html
- 5 Quanjer PH, Enright PL, Miller MR, et al. The need to change the method for defining mild airway obstruction. *Eur Respir J* 2011; 37: 720–722.
- 6 Stanojevic S, Wade A, Stocks J, et al. Reference ranges for spirometry across all ages: a new approach. *Am J Respir Crit Care Med* 2008; 177: 253–260.
- 7 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: 1324–1343.
- 8 Peto R, Speizer FE, Cochrane AL, et al. The relevance in adults of air-flow obstruction, but not of mucus hypersecretion, to mortality from chronic lung disease. Results from 20 years of prospective observation. *Am Rev Respir Dis* 1983; 128: 491–500.
- 9 Falaschetti E, Laiho J, Primatesta P, et al. Prediction equations for normal and low lung function from the Health Survey for England. *Eur Respir J* 2004; 23: 456–463.
- 10 Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general US population. *Am J Respir Crit Care Med* 1999; 159: 179–187.
- 11 Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J* 2005; 26: 319–338.
- 12 Standardization of spirometry, 1994 update. American Thoracic Society. *Am J Respir Crit Care Med* 1995; 152: 1107–1136.
- 13 Pedone C, Scarlata S, Sorino C, et al. Does mild COPD affect prognosis in the elderly? *BMC Pulm Med* 2010; 10: 35.
- 14 Vaz Fragoso CA, Concato J, McAvay G, et al. Chronic obstructive pulmonary disease in older persons: a comparison of two spirometric definitions. *Respir Med* 2010; 104: 1189–1196.
- 15 Ekberg-Aronsson M, Pehrsson K, Nilsson J-Å, et al. Mortality in GOLD stages of COPD and its dependence on symptoms of chronic bronchitis. *Respir Res* 2005; 6: 98.
- 16 Bridevaux PO, Gerbase MW, Probst-Hensch NM, et al. Long-term decline in lung function, utilisation of care and quality of life in modified GOLD stage 1 COPD. *Thorax* 2008; 63: 768–774.
- 17 Akkermans RP, Berrevoets MA, Smele IJ, et al. Lung function decline in relation to diagnostic criteria for airflow obstruction in respiratory symptomatic subjects. *BMC Pulm Med* 2012; 12: 12.
- 18 Akkermans RP, Biermans M, Robberts B, et al. COPD prognosis in relation to diagnostic criteria for airflow obstruction in smokers. *Eur Respir J* 2014; 43: 54–63.
- 19 Vaz Fragoso CA, Gill TM, McAvay G, et al. Use of lambda-mu-sigma-derived Z score for evaluating respiratory impairment in middle-aged persons. *Respir Care* 2011; 56: 1771–1777.
- 20 Vaz Fragoso CA, Concato J, McAvay G, et al. The ratio of FEV1 to FVC as a basis for establishing chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2010; 181: 446–451.
- 21 Vaz Fragoso CA, Concato J, McAvay G, et al. Respiratory impairment and COPD hospitalisation in older persons: a competing risk analysis. *Eur Respir J* 2012; 40: 37–44.
- 22 Leivseth L, Brumpton BM, Nilsen TI, et al. GOLD classifications and mortality in chronic obstructive pulmonary disease: the HUNT Study, Norway. *Thorax* 2013; 68: 914–921.
- 23 Mohamed Hoessein FAA, Zanen P, Lammers J-WJ. Lower limit of normal or FEV1/FVC <0.70 in diagnosing COPD: an evidence-based review. *Respir Med* 2011; 105: 907–915.
- 24 Mannino DM, Buist AS, Vollmer WM. Chronic obstructive pulmonary disease in the older adult: what defines abnormal lung function? *Thorax* 2007; 62: 237–241.
- 25 Mannino DM, Buist AS, Vollmer WM. Authors' reply. *Thorax* 2007; 62: 1108–1109.
- 26 Casanova C, Cote C, de Torres JP, et al. Inspiratory-to-total lung capacity ratio predicts mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2005; 171: 591–597.
- 27 Cummings SR, Bates D, Black DM. Clinical use of bone densitometry: scientific review. *JAMA* 2002; 288: 1889–1897.
- 28 World Health Organization. World Health Organization Child Growth Standards. Geneva, WHO Press, 2006. Available from: www.who.int/childgrowth/standards/Technical_report.pdf
- 29 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.
- 30 Barnes PJ, Celli BR. Systemic manifestations and comorbidities of COPD. *Eur Respir J* 2009; 33: 1165–1185.
- 31 Lange P, Marott JL, Vestbo J, et al. Prediction of the clinical course of chronic obstructive pulmonary disease, using the new GOLD classification: a study of the general population. *Am J Respir Crit Care Med* 2012; 186: 975–981.
- 32 Johannessen A, Nilsen RM, Storebø M, et al. Comparison of 2011 and 2007 Global Initiative for Chronic Obstructive Lung Disease guidelines for predicting mortality and hospitalization. *Am J Respir Crit Care Med* 2013; 188: 51–59.
- 33 Fingleton J, Weatherall M, Beasley R. Towards individualised treatment in COPD. *Thorax* 2011; 66: 363–364.