

GOLD 2011 disease severity classification in COPDGene: a prospective cohort study



MeiLan K Han, Hana Muellerova, Douglas Curran-Everett, Mark T Dransfield, George R Washko, Elizabeth A Regan, Russell P Bowler, Terri H Beaty, John E Hokanson, David A Lynch, Paul W Jones, Antonio Anzueto, Fernando J Martinez, James D Crapo, Edwin K Silverman, Barry J Make

Summary

Background The 2011 GOLD (Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease [COPD]) consensus report uses symptoms, exacerbation history, and forced expiratory volume (FEV₁)% to categorise patients according to disease severity and guide treatment. We aimed to assess both the influence of symptom instrument choice on patient category assignment and prospective exacerbation risk by category.

Methods Patients were recruited from 21 centres in the USA, as part of the COPDGene study. Eligible patients were aged 45–80 years, had smoked for 10 pack-years or more, and had an FEV₁/forced vital capacity (FVC) <0.7. Categories were defined with the modified Medical Research Council (mMRC) dyspnoea scale (score 0–1 vs ≥2) and the St George's Respiratory Questionnaire (SGRQ; ≥25 vs <25 as a surrogate for the COPD Assessment Test [CAT] ≥10 vs <10) in addition to COPD exacerbations in the previous year (<2 vs ≥2), and lung function (FEV₁% predicted ≥50 vs <50). Statistical comparisons were done with *k*-sample permutation tests. This study cohort is registered with ClinicalTrials.gov, number NCT00608764.

Findings 4484 patients with COPD were included in this analysis. Category assignment using the mMRC scale versus SGRQ were similar but not identical. On the basis of the mMRC scale, 1507 (33.6%) patients were assigned to category A, 919 (20.5%) to category B, 355 (7.9%) to category C, and 1703 (38.0%) to category D; on the basis of the SGRQ, 1317 (29.4%) patients were assigned to category A, 1109 (24.7%) to category B, 221 (4.9%) to category C, and 1837 (41.0%) to category D (κ coefficient for agreement, 0.77). Significant heterogeneity in prospective exacerbation rates (exacerbations/person-years) were seen, especially in the D subcategories, depending on the risk factor that determined category assignment (lung function only [0.89, 95% CI 0.78–1.00]), previous exacerbation history only [1.34, 1.0–1.6], or both [1.86, 1.6–2.1; *p*<0.0001]).

Interpretation The GOLD classification emphasises the importance of symptoms and exacerbation risk when assessing COPD severity. The choice of symptom measure influences category assignment. The relative number of patients with low symptoms and high risk for exacerbations (category C) is low. Differences in exacerbation rates for patients in the highest risk category D were seen depending on whether risk was based on lung function, exacerbation history, or both.

Funding National Heart, Lung, and Blood Institute, and the COPD Foundation through contributions from AstraZeneca, Boehringer Ingelheim, Novartis, and Sepracor.

Introduction

The development of a disease staging system in Chronic Obstructive Pulmonary Disease (COPD) to assess severity and determine treatment algorithms has proven challenging. The Global initiative for chronic Obstructive Lung Disease (GOLD) 2006 staging system used forced expiratory volume (FEV₁) to determine disease severity.¹ In large population studies, lung function correlated reasonably well with many disease outcomes but was poorly predictive of dyspnoea, quality of life, and exacerbation frequency.^{2,3} Lung function alone does not completely capture the heterogeneity that exists among patients with COPD.⁴ Hence, the GOLD 2011 consensus report proposed a new classification system for COPD to more comprehensively assess disease severity,⁵ incorporating symptoms with either a dyspnoea measure (the modified Medical Research Council [mMRC] dyspnoea score) or a health status measure (the

COPD Assessment Test [CAT] score) in addition to COPD exacerbation history and airflow limitation measured by FEV₁ (figure 1). We aimed to use the COPDGene cohort to establish whether the choice of a symptom versus health-status measure significantly influences baseline category assignment and whether the categories differ, especially with respect to exacerbation risk.

Methods

Study design and patients

The COPDGene Study is a multicentre investigation of the genetic epidemiology of smoking-related lung disease. Patients were recruited at 21 academic clinical centres in the USA. However, participants were recruited through various mechanisms including general public advertising and screening of primary-care and pulmonary clinics. Patients were eligible if they were aged 45–80 years

Published Online
August 18, 2012
[http://dx.doi.org/10.1016/S2213-2600\(12\)70044-9](http://dx.doi.org/10.1016/S2213-2600(12)70044-9)

See Online/Comment
[http://dx.doi.org/10.1016/S2213-2600\(12\)70045-0](http://dx.doi.org/10.1016/S2213-2600(12)70045-0)

Division of Pulmonary and Critical Care, University of Michigan, Ann Arbor, MI, USA (M K Han MD, Prof F J Martinez PhD); Epidemiology, GlaxoSmithKline R&D, Stockley Park, Middlesex, UK (H Muellerova PhD); Division of Biostatistics and Bioinformatics (Prof D Curran-Everett PhD), Department of Medicine (E A Regan MD, R P Bowler MD, Prof J D Crapo MD, Prof B J Make MD), Department of Radiology (Prof D A Lynch MD), National Jewish Health, Denver, CO, USA; Division of Pulmonary and Critical Care, University of Alabama Birmingham, Birmingham, AL, USA (M T Dransfield MD); Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA (G R Washko MD); Johns Hopkins School of Public Health, Baltimore, MD, USA (Prof T H Beaty); Department of Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA (Prof J E Hokanson PhD); Division of Clinical Science, St George's, University of London, London, UK (Prof P W Jones PhD); Pulmonary/Critical Care Medicine, University of Texas Health Science Center and South Texas Veterans Healthcare System, San Antonio, TX, USA (Prof A Anzueto MD); and Channing Division of Network Medicine and Pulmonary and Critical Care Division, Brigham and Women's Hospital, Boston, MA, USA (E K Silverman MD)

Correspondence to: Dr MeiLan K Han, 3916 Taubman Center, 1500 East Medical Center Drive, Ann Arbor, MI 48109, USA mrking@umich.edu

For more on the COPDGene Study see <http://www.copdgene.org>

and had smoked 10 pack-years or more. Patients must also have had the ability to undertake spirometry and not have had an exacerbation in the 4 weeks before enrolment. A full list of inclusion and exclusion criteria has been described previously.⁶ Patients were enrolled from Jan 10, 2008, to April 15, 2011. Patients included in this analysis were part of the patient dataset obtained on April 16, 2012, comprising a total of 10 300 patients. This dataset included

those with a post-bronchodilator FEV₁/forced vital capacity (FVC) ratio ≤ 0.7 . Entry criteria for the COPDGene study were based on smoking history not COPD status. 5816 patients did not have airflow obstruction as defined by FEV₁/FVC < 0.70 and therefore did not qualify as having COPD by GOLD criteria. All 4484 with COPD were included in this analysis; a full baseline visit was required to determine inclusion in the 10 300 cohort. 3723 (83%) of 4484 patients with COPD completed additional longitudinal follow-up. All participants provided written informed consent. The research protocol was approved by the ethics and review boards of the participating centres.

Procedures

We used self-administered questionnaires to record demographic and medical history data. Symptoms were quantified with both the mMRC⁷ and the St George's Respiratory Questionnaire (SGRQ).⁸ The mMRC scale 0–4 was developed by the American Thoracic Society as a modification of the originally proposed British Medical Research Council dyspnoea index (scale 1–5).⁹ We determined the distribution of patients in GOLD categories independently with the mMRC (dyspnoea) first and then with the SGRQ (health status). The GOLD 2011 classification stratifies first on the basis of symptoms with either dyspnoea (mMRC 0–1 vs ≥ 2) or health status (CAT < 10 or ≥ 10) score resulting in two low-symptom categories (A and C) and two high-symptom categories (B and D). While the CAT was not used in the COPDGene study, it has been previously shown that a CAT score of 10 (scale 0–40) is comparable with an SGRQ score of 25 (scale 0–100); available data show 31% of CAT scores will differ from the equivalent SGRQ score by $\leq 5\%$, 60% will differ by $\leq 10\%$, and 90% will differ by $\leq 20\%$.^{10,11} Hence, an SGRQ score of 25 was used as a substitute measure for a CAT score of 10. Further outlined in the GOLD 2011 report, exacerbation risk is assessed with either airflow limitation measured by FEV₁% predicted ($< 50\%$ or $\geq 50\%$), or COPD exacerbation history (0–1 vs ≥ 2) in the previous year to stratify patients into low-risk categories (A and B) versus high-risk (C and D) categories (figure 1). Additionally, the GOLD 2011 report outlines that where exacerbation risk, as determined by FEV₁ or previous exacerbation history, is not identical, risk should be defined by the method showing the higher risk.⁵ For this analysis, we have further stratified these patients. In addition to meeting symptom criteria for their respective categories, patients in C1 and D1 categories meet FEV₁ criteria only for C and D, respectively; C2 and D2 meet exacerbation criteria only; C3 and D3 meet both exacerbation and FEV₁ criteria.

Exacerbations were defined by use of antibiotics, steroids, or both or admission to the hospital for a respiratory flare-up on the American Thoracic Society Respiratory Disease questionnaire.¹² Severe exacerbations

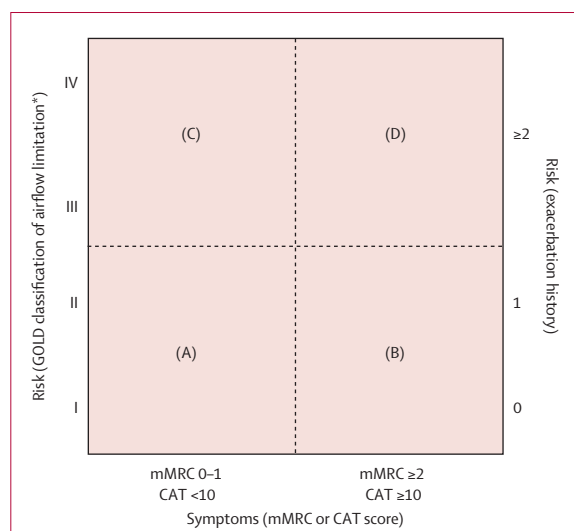


Figure 1: GOLD 2011 classification system
 Reproduced with permission from GOLD.⁵ When assessing exacerbation risk, choose the highest risk according to GOLD grade or exacerbation history. GOLD I: FEV₁ $\geq 80\%$ predicted. GOLD II: FEV₁ 50–79% predicted. GOLD III: FEV₁ 30–49% predicted. GOLD IV: FEV₁ $< 30\%$ predicted. GOLD=Global initiative for chronic Obstructive Lung Disease. mMRC=modified Medical Research Council dyspnoea scale. CAT=Chronic obstructive pulmonary disease Assessment Test. *GOLD I–IV classification of airflow limitation based on FEV₁ from GOLD executive summary 2006.¹

	mMRC classification system		CAT (SGRQ) classification method	
	mMRC 0–1	mMRC ≥ 2	SGRQ < 25	SGRQ > 25
Symptom category				
A	A (33.6% [1507])	..	A (29.4% [1317])	..
B	..	B (20.5% [919])	..	B (24.7% [1109])
C	C (7.9% [355])	..	C (4.9% [221])	..
D	..	D (38.0% [1703])	..	D (41.0% [1837])
Symptom subcategories				
C1*	C1 (5.8% [259])	..	C1 (3.9% [173])	..
C2†	C2 (1.5% [68])	..	C2 (0.8% [38])	..
C3‡	C3 (0.6% [28])	..	C3 (0.2% [10])	..
D1*	..	D1 (24.4% [1096])	..	D1 (26.4% [1182])
D2†	..	D2 (5.0% [222])	..	D2 (5.6% [252])
D3‡	..	D3 (8.6% [385])	..	D3 (9.0% [403])

N=4484. CAT=COPD Assessment Test. SGRQ=St George's Respiratory Questionnaire. mMRC=modified Medical Research Council. FEV₁=forced expiratory volume. COPD=chronic obstructive pulmonary disease. GOLD=Global initiative for chronic Obstructive Lung Disease. *Meets FEV₁ criteria ($< 50\%$ FEV₁ predicted). †Meets exacerbation criteria (≥ 2 COPD exacerbations in past 12 months). ‡Meets exacerbation and FEV₁ criteria ($< 50\%$ FEV₁ predicted and ≥ 2 COPD exacerbations in past 12 months).

Table 1: Distribution of patients into GOLD 2011 categories by symptom measure (mMRC vs SGRQ)

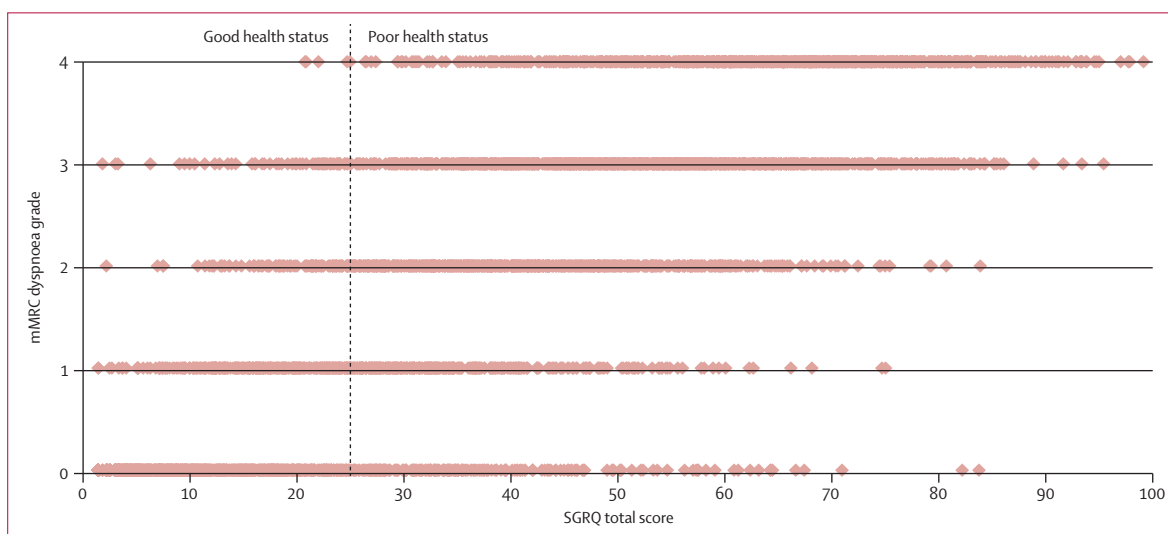


Figure 2: Distribution of the SGRQ total score by the mMRC dyspnoea grade in the COPDGene cohort
mMRC=modified Medical Research Council. SGRQ=St George's Respiratory Questionnaire. Datapoints represent individual patients.

were defined as those requiring assessment in the emergency department or hospital admission. Data about exacerbation history in the previous year were gathered at baseline. Prospective exacerbation data were gathered through a longitudinal follow-up protocol done every 6 months by an automated telephonic or web-based inquiry. Patients not reached by the automated system were contacted by a research coordinator (appendix).

Prospective exacerbation rates were calculated for all patients with COPD where longitudinal follow-up data were available. Length of follow-up varied on the basis of time of enrolment with a mean follow-up period of 20 months (SD 11). April 22, 2012, was used as the cutoff date for available longitudinal data. Exacerbation rate during follow-up was calculated by dividing total exacerbations by the number of years of follow-up. For patients who had been followed up for only 6 months, exacerbation rate was calculated by doubling the number of exacerbations that occurred within that 6 months.

Patients underwent standardised spirometry before and after the administration of 180 µg of inhaled albuterol. Independent review of spirometric tracings was done to ensure quality control of spirometry data.¹³ All patients underwent a standard 6 min walk distance test. BODE (the body-mass index, airflow obstruction, dyspnoea, and exercise) index was calculated according to previously described methods.¹⁴

Statistical analyses

All data are presented as mean (SD) where appropriate. Analyses were done with SAS (version 9.2). Statistical comparisons were done with *k*-sample permutation tests using R (version 2.15.0). 95% CIs were estimated by traditional normal-theory method and verified by 10 000 bootstrap replications.

This trial is registered with ClinicalTrials.gov, number NCT00608764.

Role of the funding source

The sponsor of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

See Online for appendix

Results

Of the 4484 patients with COPD as classified with the mMRC, the distribution of symptom categories A–D was 33.6% (1507; A), 20.5% (919; B), 7.9% (355; C), and 38.0% (1703; D). Using the SGRQ, the distribution was similar (29.4% [1317; A], 24.7% [1109; B], 4.9% [221; C], and 41.0% [1837; D]; table 1, appendix). The κ coefficient of agreement for patients' classification by the two symptom measures was 0.77, suggesting good but not identical agreement. Some patients reported no (0) or mild (1) dyspnoea but also reported very poor health status (high SGRQ scores; figure 2). The mean mMRC score of 2 corresponded with a mean SGRQ score of 39 (SD 14), whereas a mean mMRC score of 1 corresponded to a mean SGRQ score of 26 (13; table 2).

The largest area of disagreement in reassignment of patients and GOLD classification seem to be in group C, where 166 (46.7%) of patients in group C according to the mMRC would have been classified identically by the SGRQ; 166 (75.1%) in group C according to the SGRQ would have been classified identically by the mMRC (figure 3, appendix). Few patients were classified as C, irrespective of the symptom measure chosen. Additionally, most patients classed as C and D met FEV₁ criteria alone (C1 and D1) as opposed to exacerbation

criteria (C2 and D2), or the combination of exacerbation and FEV₁ criteria (C3 and D3).

Further data analyses used the SGRQ score to assign symptom severity within the GOLD classification as opposed to mMRC, because available evidence suggests both the SGRQ and CAT are more repeatable and more sensitive to change than the mMRC.¹⁵⁻¹⁹ At baseline, ages were roughly similar across groups, current smoking was less prevalent in high-risk groups, and BODE scores were especially high for the D1 and D3 subgroups (table 3). Compared with patients in the GOLD category B, patients in category A had better lung function (p<0.0001), lower symptom scores (p<0.0001), and had lower exacerbation rates in longitudinal follow-up (p<0.0001; table 4).

While SGRQ scores were similar within the C and D subcategories, significant variability exists within these subcategories. While exacerbation rates were not statistically different for C subcategories, patients in the D subcategory did have significantly different prospective exacerbation rates (p<0.0001). Exacerbation rates were highest in the D3 group (met exacerbation

and FEV₁ criteria) at 1.86 exacerbations/person-year, followed by D2 (met exacerbation criteria) at 1.34 exacerbations/person-year and D1 (met FEV₁ criteria) at 0.89 exacerbations/person-year. Severe exacerbation rates followed a similar pattern (p<0.0001). Lung function also differed significantly within this group (p<0.0001; table 4). In terms of exacerbation frequency, no significant difference in prospectively assessed total exacerbation rates were seen between group B or any of the C subgroups (p=0.35) when compared with a four-sample permutation test. Similarly, no difference in severe exacerbation rates between group B and any of the C subgroups (p=0.23) was seen. However, the small number of patients in the C groups and relatively wide 95% CIs limit our power to compare the C subgroups.

The percentage of patients on any recommended drug regimen increased with GOLD category severity (appendix). Use of long-acting muscarinic antagonists (LAMAs) alone exceeded use of long-acting β agonists (LABAs) alone for patients in group A and B (87 [6.6%] patients in group A and 301 (27.1%) in group B on LAMA vs 14 [1.1%] in group A and 73 [6.6%] on LABA in group B). The combined use of LAMAs and LABAs as recommended for patients in groups C and D was generally low (six [2.7%] of 221 in group C and 144 [7.8%] of 1837 in group D). Inhaled corticosteroid and LABA was the most common combination therapy. The use of an inhaled corticosteroid and LAMA regimen as recommended for patients in groups C and D was also low (none in group C and 170 [9.3%] of 1837 in group D). Combination therapy with inhaled corticosteroid, LABA, and LAMA was not used in group C and was given to 802 (43.7%) of 1837 patients in group D. Imaging data show greater extent of emphysema and more gas trapping in higher risk groups in general than in lower risk groups (appendix).

	SGRQ total score	n (%)
0	14 (13)	1227 (27%)
1	26 (13)	622 (14%)
2	39 (14)	725 (16%)
3	50 (16)	1156 (26%)
4	63 (15)	740 (17%)

Data are mean (SD) or number (%). N=4484. SGRQ=St George's Respiratory Questionnaire. mMRC=modified Medical Research Council.

Table 2: Distribution of baseline SGRQ total score by mMRC grade

		mMRC assignment							
		A	B	C1	C2	C3	D1	D2	D3
SGRQ assignment	A	921	396						
	B	137	972						
	C1			80			91		
	C2				15			23	
	C3					7			3
	D1			61			1121		
	D2				15			237	
	D3					4			399

Figure 3: Reassignment of GOLD categories by symptom measure (mMRC vs SGRQ)
 GOLD category explanations: (A) low symptoms and low risk; (B) high symptoms and low risk; (C) low symptoms and high risk; and (D) high symptoms and high risk. Groups C1-C3 and D1-D3 were classified on the basis of the qualifying criteria: patients in C1 or D1 qualified on the basis of airflow limitation, FEV₁ <50% predicted, only; C2 or D2 qualified on the basis of frequent history of COPD exacerbation (≥2 events per past 12 months); and C3 or D3 qualified on the basis of both airflow limitation and history of frequent COPD exacerbation.

Discussion

The GOLD 2011 consensus report outlines a new classification system for COPD combining spirometry, symptom assessment, and exacerbation risk to identify disease severity and appropriate therapeutic regimens. This approach uses available evidence showing that current therapies can improve lung function and reduce symptoms and exacerbation frequency.^{20,21} Gathering data to assess the validity of this method is important (panel). We postulated that analysis of the COPDGene cohort would provide insight into the practical application of this approach.

We showed that choice of symptom measure, dyspnoea (mMRC) versus health status (SGRQ as a surrogate for the CAT), can alter category assignment; the relative number of patients with low symptoms and high risk (category C) is small; the apparent risk for exacerbations in category C also does not seem to be substantially different from category B (high symptoms with low risk

	A	B	C1	C2	C3	D1	D2	D3
Number of patients	1317 (29%)	1109 (25%)	173 (4%)	38 (1%)	10 (<1%)	1182 (26%)	252 (6%)	403 (9%)
Age (years)	63 (9; 62–63)	62 (9; 61–62)	66 (8; 65–67)	65 (8; 62–68)*	65 (5; 62–69)*	65 (8; 64–65)	61 (9; 60–62)	63 (8; 62–63)
Sex (male)	780 (59%)	582 (53%)	103 (60%)	12 (32%)	6 (60%)	703 (59%)	111 (44%)	212 (53%)
Smoking, pack-years	45 (23; 44–46)	54 (30; 52–55)	50 (24; 47–54)	44 (19; 37–50)*	62 (22; 46–77)*	57 (29; 55–58)	51 (27; 48–58)	55 (27; 52–57)
Current smoking	664 (50%)	608 (55%)	58 (34%)	8 (21%)	1 (10%)	381 (32%)	112 (28%)	112 (28%)
Body-mass index (kg/m ²)	27 (5; 27–28)	29 (6; 29–29)	27 (6; 26–27)	28 (5; 27–30)*	25 (5; 22–29)*	27 (6; 27–28)	30 (7; 29–31)	27 (7; 27–28)
BODE index	0.4 (0.7; 0.4–0.5)	1.9 (1.3; 1.8–2.0)	2.7 (1.0; 2.5–2.8)	0.8 (1.1; 0.4–1.1)*	3.0 (1.2; 2.1–3.9)*	4.6 (1.4; 4.5–4.7)	2.5 (1.2; 2.3–2.6)	5.1 (1.3; 5.0–5.2)

Data are number (%) or number (SD; 95% CI). *Because of small sample size, the 95% CI might be an unreliable estimate; this is reflected in the relative width of this 95% CI compared with the 95% CIs in other groups. BODE=body-mass index, airflow obstruction, dyspnoea, and exercise. GOLD=Global initiative for chronic Obstructive Lung Disease.

Table 3: Baseline characteristics by GOLD risk groups (categorised using SGRQ)

	A	B	C1	C2	C3	D1	D2	D3
Number of patients	1317 (29%)	1109 (25%)	173 (4%)	38 (1%)	10 (<1%)	1182 (26%)	252 (6%)	403 (9%)
Lung function and exercise capacity								
FEV ₁ , % predicted	78% (15; 77–79)	68% (13; 67–79)	41% (7; 40–42)	68% (12; 64–72)*	40% (7; 35–44)*	41% (7; 33–34)	65% (11; 64–66)	32% (10; 31–33)
FEV ₁ /FVC	0.62 (0.1; 0.62–0.62)	0.59 (0.1; 0.58–0.59)	0.45 (0.1; 0.43–0.46)	0.59 (0.1; 0.56–0.61)*	0.39 (0.1; 0.35–0.44)*	0.40 (0.1; 0.39–0.40)	0.58 (0.1; 0.57–0.59)	0.38 (0.1; 0.37–0.39)
Positive bronchodilator response (>12% or >200 cm ³)	409 (31%)	392 (35%)	64 (37%)	12 (32%)	3 (30%)	426 (36%)	86 (34%)	153 (38%)
6-min walk distance (m)	459 (99; 453–464)	374 (111; 368–381)	400 (95; 386–415)	433 (85; 405–461)*	387 (79; 331–443)*	308 (112; 302–315)	363 (108; 349–377)	289 (114; 277–300)
Symptoms								
SGRQ, total	11 (8; 10–11)	45 (15; 44–46)	17 (6; 16–18)	17 (7; 14–19)*	18 (5; 14–21)*	51 (16; 50–52)	54 (16; 52–56)	60 (15; 58–61)
Respiratory symptoms								
mMRC	0.4 (0.8; 0.4–0.5)	2.2 (1.2; 2.1–2.3)	0.9 (1.0; 0.7–1.0)	0.9 (1.0; 0.6–1.2)*	0.7 (1.3; 0.1–6)*	2.8 (1.1; 2.8–2.9)	2.7 (1.1; 2.6–2.9)	3.2 (0.9; 3.1–3.3)
Chronic bronchitis	137 (10%)	389 (35%)	25 (14%)	10 (26%)	2 (20%)	333 (28%)	112 (44%)	156 (39%)
Total exacerbation rate in the year before enrolment (events/year)	0.07 (0.26; 0.06–0.08)	0.27 (0.44; 0.24–0.29)	0.18 (0.39; 0.13–0.24)	2.45 (0.80; 2.19–2.71)*	2.00 (0.71; 2.00–3.01)*	0.35 (0.48; 0.32–0.38)	2.90 (1.26; 2.74–3.05)	3.10 (1.36; 2.96–3.23)
Prospective total annual exacerbation rate (events/year)	0.17 (0.60; 0.13–0.20)	0.55 (1.4; 0.46–0.64)	0.58 (1.7; 0.31–0.84)*	0.52 (1.2; 0.08–0.97)*	1.39 (2.1; 0.0–3.0)*	0.89 (1.9; 0.78–1.0)	1.34 (2.2; 1.0–1.6)*	1.86 (2.5; 1.6–2.1)
Prospective annual severe exacerbation rate (events/year)	0.02 (0.12; 0.01–0.03)	0.13 (0.44; 0.10–0.16)	0.06 (0.20; 0.03–0.09)*	0.13 (0.52; 0.0–0.31)*	0.11 (0.33; 0.0–0.37)*	0.27 (0.79; 0.22–0.32)	0.57 (1.6; 0.34–0.79)*	0.62 (1.2; 0.49–0.75)

Data are number (%) or number (SD; 95% CI). Longitudinal follow-up data were available in 3723 (83%) of 4484 patients. *Because of small sample size, the 95% CI might be an unreliable estimate; this is reflected in the relative width of this 95% CI compared with the 95% CIs in other groups. FEV₁=forced expiratory volume. SGRQ=St George's Respiratory Questionnaire. mMRC=modified Medical Research Council. GOLD=Global initiative for chronic Obstructive Lung Disease.

Table 4: Symptoms and exacerbation frequency according to GOLD risk groups (categorised using SGRQ)

for exacerbations), although the relatively small number of patients might limit our power to detect differences; and for patients in category D with the greatest symptoms and highest risk for exacerbations, the reason for category assignment (lung function and exacerbation history versus lung function alone versus exacerbation history alone) significantly alters future exacerbation risk.

In the comparison of mMRC with SGRQ, a very wide range of SGRQ scores was seen at every mMRC level. The SGRQ and CAT are multidimensional methods, which assess not only dyspnoea but also cough, sputum production, fatigue, and the effect of these symptoms on activities and daily life¹⁵ such that imperfect correlation between them and the mMRC is not surprising. Since the category assignments produced by each symptom

measure are not identical, a potential refinement of the GOLD classification schema would be to choose one measure only. If two measures are retained, however, another possibility would be to move the cutpoint for mMRC from 2 to 1 because an mMRC of 2 corresponds with an SGRQ of 39 whereas an mMRC of 1 corresponds to an SGRQ of 26 (table 2). An SGRQ score of 25 corresponds with a CAT score of 10. An mMRC score of 2 or CAT score of 10 are the currently recommended GOLD 2011 symptom stratification measures. In practice, an advantage of the mMRC is its brevity, but it only addresses disability due to breathlessness. However, the CAT, which was also proposed by GOLD and strongly correlates with SGRQ, has broader coverage of the effect of COPD on patient health. The CAT has good

Panel: Research in context**Systematic review**

The GOLD 2011 classification system recommends treatment algorithms on the basis of an assessment of symptoms and risk using thresholds for symptom scores, lung function, and exacerbation history. However, no prospective assessment has previously been done to determine whether the choice of symptom measure matters, how many patients with COPD actually meet criteria for each of the subgroups, and whether the subgroups actually differ in symptoms and risk for exacerbations. We searched PubMed between July 1, 2011, and July 30, 2012, to identify prospective studies that validate the GOLD 2011 criteria, without language restrictions. We used the following search term: "GOLD COPD 2011". On the basis of our review of the literature, we did not identify any available published data that prospectively validated the GOLD 2011 criteria, because this stratification system was recently introduced. Therefore, at this time, it is difficult to compare our results to any other similar analyses from other cohorts.

Interpretation

From a practical standpoint, our data suggest that classification assignment does depend on the symptom metric used, patients with low symptoms but high risk for exacerbation are less common, and significant variability in risk exists even among high-risk individuals such that clinicians should be wary of patients who have low lung function, high symptoms, and previous history of exacerbation (group D3), because these individuals are at greatest risk.

repeatability¹⁰ and is responsive both to pulmonary rehabilitation^{22,23} as well as exacerbation onset and recovery,^{23,24} suggesting that the CAT might be better than the mMRC.

Another interesting finding was that category C was relatively small, suggesting it is unusual for patients who are at high risk for exacerbations to also not report significant symptoms. Even for the few patients in category C, we noted that the exacerbation frequency of this low-symptom and high-risk group is no different from category B, the high-symptom and low-risk group. These data would suggest from a therapeutic standpoint that symptoms differentiate the B and C categories.

In general, patients in group D had the most exacerbations and severe exacerbations. However, among these patients, significant heterogeneity exists between D1, D2, and D3 on the basis of the reason for assignment to that category: exacerbation history and FEV₁% predicted, exacerbation history alone, or airflow limitation alone. The small number of patients in group C limits our power to compare differences between C1, C2, and C3 subcategories. Total exacerbation frequency and severe exacerbation frequency were highest for patients in D3 who met both exacerbation history and lung-function risk factors versus either risk factor alone. In practice,

more patients in group D3 were given inhaled corticosteroids, LABA, and LAMA therapies than in D1 or D2, suggesting that higher exacerbation rates in patients in D3 were not related to less intense therapy and clinicians were already more aggressive in treating patients in this category.

The D1 subcategory defined by FEV₁% predicted is by far the largest subcategory, suggesting that treatment recommendations for the D category should aim to target these patients. Patients in the D2 group, while fewer in number than in D1, notably had much higher lung function, suggesting mechanisms for exacerbations could differ in this subcategory.

We acknowledge limitations to our analysis. The COPDGene cohort is not a true population-based sample and therefore might not represent the true distribution of COPD severity in the general population. The inclusion or exclusion criteria were primarily focused on concomitant respiratory disorders, life-threatening disorders, and previous chest surgery. The inability to undertake spirometry and a history of exacerbation in the 4 weeks before enrolment were additional exclusion criteria that could have biased the population against sicker patients. Reported therapies could have been biased according to prescribing patterns at academic centres. However, patients were recruited from multiple centres in the USA not only from within the participating clinical centres but also from general community advertising. Furthermore, patients were roughly split into groups with about half of the cohort classed as GOLD I–II and the other half GOLD III–IV, which makes this cohort ideal for examining the GOLD 2011 classification schema. The COPDGene study is also one of the largest COPD cohorts with the detailed type of information available needed to stratify patients via the GOLD 2011 classification system and examine prospective outcomes. Another advantage of this cohort is its relatively recent recruitment reflecting current prescribing practices. The SGRQ and CAT, however, are not identical although the CAT and SGRQ show a very stable relation across the scaling range¹¹ and a high degree of correlation ($r=0.84$) in a primary-care population of patients with COPD in seven European countries.¹⁰ We also acknowledge that the GOLD 2011 classification system is intended to assess risk for exacerbations, hospital admissions, and death. We included the longitudinal data for total number of exacerbations and exacerbations that required hospital admission but not mortality data, which was not available at the time of this analysis. Additionally, prospective data on exacerbation rates was available in 3723 (83%) patients, which could result in biased estimations of actual exacerbation rates. Sicker patients, for instance, might be in the hospital and not at home to do either the web or telephone based surveys. 193 (5.2%) of 3723 patients had 6 months of prospective data or less

and the annual exacerbation rate was estimated on the basis of available data. This estimation is a potential source of bias because such estimates could be too high or too low if data were gathered primarily in the winter when exacerbation rates are higher or in the summer when exacerbation rates are lower.

In summary, the GOLD 2011 classification system identifies COPD subcategories that differ in disease severity assessed by lung function, symptoms, and exacerbation frequency. Our analysis, however, showed the difficulty of using more than one risk stratifier because exacerbation history and FEV₁ do not behave identically in predicting risk. FEV₁ is also problematic for risk stratification because it relates both to symptoms and risk. As a result, we noted that patients in group C with low symptoms as defined by SGRQ but high risk as defined by FEV₁ history of exacerbation, or both are a relatively small group of individuals. Furthermore, significant variability in exacerbation rates in the high-risk groups occurred depending on whether risk stratification was based on FEV₁% predicted, history of exacerbation, or both. Group D could be further subdivided as we have done in this analysis, because patients in the D3 group had the highest exacerbation rates and were already being treated the most intensely.

We also showed that the choice of symptom measure can significantly alter category assignment. GOLD currently stratifies patients on the basis of an mMRC of 2 or CAT 10 (SGRQ 25). In our analysis, an mMRC of 2 corresponds roughly to an SGRQ of 39 whereas a cutoff of 1 would correspond with an SGRQ of 26, much closer to the equivalent CAT score of 10. Moving the mMRC cutpoint to 1 could improve the likelihood that patients would be assigned to the same risk group regardless of symptom metric chosen. However, irrespective of the cutpoint, patients identified by the two symptom metrics will not be identical (figure 2). The GOLD stratification system will inevitably be used to identify patients with similar symptom and risk profiles for the purposes of designing clinical trials and assessing aggregate data on health care outcomes to examine the efficacy of treatments. Care should be taken when comparing patient populations identified with one symptom metric versus another. Additionally, noise within any one metric is expected, and inherent to any classification system with three axes and more than one choice of metric per axis; the potential for variability in patient assignment increases accordingly. Therefore, further prospective investigation will be needed to assess both short-term test-retest patient classification and longer-term patient classification stability. However, when applied in clinical practice, a substantial strength of the GOLD classification system is that it will challenge health care providers to think about their patients with COPD in a more multidimensional way. Importantly, we showed that COPD populations indeed are very heterogeneous with poor correlation between lung

function, symptoms, and exacerbation frequency, confirming the need for a classification system based on more than lung function alone.

Contributors

MKH, MTD, GRW, RPB, DAL, AA, FJM, JDC, EKS, and BJM aided with data collection. MKH and DC-E analysed the data. MKH, HM, DC-E, MTD, GRW, EAR, RPB, THB, JEH, DAL, PWJ, AA, FJM, JDC, EKS, and BJM contributed to data interpretation and preparation of the report.

Conflicts of interest

MKH participated in advisory boards for Boehringer Ingelheim, Pfizer, GlaxoSmithKline, Genentech, Novartis, and Medimmune; participated on speaker's bureaus for Boehringer Ingelheim, Pfizer, GlaxoSmithKline, Grifols therapeutics, and the National Association for Continuing Education, and WebMD; has consulted for Novartis and United Biosource Corporation; and has received royalties from UpToDate and ePocrates. HM is an employee of GlaxoSmithKline R&D and own shares and stock options of GlaxoSmithKline. MTD has consulted for GlaxoSmithKline and Boehringer Ingelheim; his institution has received funds to undertake clinical research trials for NHLBI, Aeirs, GlaxoSmithKline, Boehringer Ingelheim, Pfizer, Otsuka, Boston Scientific, and Contocor. GRW has consulted for Spiration. DAL has served as consultant to several trials in idiopathic pulmonary fibrosis, sponsored by Gilead, Actelion, Johnson and Johnson, and Intermune; is a consultant for Perceptive Imaging; and receives research support from Siemens. PWJ consulted for Almirall, AstraZeneca, Bayer, Forest, GlaxoSmithKline, Novartis, Pearl, Roche, and Spiration; has received grants through his institution from GlaxoSmithKline; and has served on speakers' bureau for Almirall, GlaxoSmithKline, and Novartis; and has received no royalties. AA is a member of GOLD scientific and Executive committees; has participated in scientific meetings or courses organised and financed by various pharmaceutical companies including Boehringer Ingelheim, Bayer Healthcare, GlaxoSmithKline, and Forest Laboratories; has been a consultant for AstraZeneca, Boehringer Ingelheim, Pfizer, GlaxoSmithKline, Bayer Healthcare, Forest Laboratories, Intermune, and Amgen; has been the principal investigator for research done at his institution (University of Texas Health Science Center at San Antonio) who were paid for participating in multicentre clinical trials sponsored by GlaxoSmithKline, Bayer-Schering Pharma, Lilly, National Institutes of Health, and COPDGene. FJM has served on advisory boards relating to COPD-related topics for GlaxoSmithKline, MedImmune, AstraZeneca, Merck, Pearl Therapeutics, Novartis, United BioSource Corporation, Forest Laboratories, and Almirall; he has consulted for Actelion Pharmaceuticals, Boehringer Ingelheim, Nycomed, Forest Laboratories, F. Hoffmann-La Roche; Bayer, Merck/Schering-Plough, Health Learning Systems, Talecris Biotherapeutics, Comgenex, fb Communications, BoomComm, and Actelion; has served on speaker's bureaus for GlaxoSmithKline, National Association for Continuing Education, Med-Ed, Potomac Center for Medical Education, Pfizer, Boehringer Ingelheim, Merck/Schering-Plough, Vox Medica, American Lung Association, WebMD, ePocrates, AstraZeneca, France Foundation, CME Incite, and Altana/Nycomed; his institution has received funds from Boehringer Ingelheim for a clinical trial; has received royalties from Associates in Medical Marketing and Castle Connolly; has developed educational materials for the France Foundation, HIT Global, and ePocrates; has served on steering committees for clinical trials supported by GlaxoSmithKline, Nycomed, Forest Laboratories, and Actelion. EKS has received grant support and consulting fees from GlaxoSmithKline for studies of COPD genetics; has received honoraria and consulting fees from AstraZeneca; and has received consulting fees from Merck. BJM has participated in advisory boards, speaker's bureaus, consultations, and multicentre clinical trials with funding from the National Heart, Lung, and Blood Institute, Abbott Laboratories, Astellas Pharma, AstraZeneca, Boehringer Ingelheim, Dey Pharma, Embryon, Forest Laboratories, GlaxoSmithKline, Ikaria, MedImmune, Merck, Nabi Biopharmaceuticals, Nycomed, Novartis, Pfizer, Respironics, Merck/Schering-Plough, SeQual Technologies, and Talecris Biotherapeutics. DC-E, EAR, RPD, THB, JEH, and JDC declare no conflicts of interest.

Acknowledgments

The COPDGene project is supported by grants from the National Heart, Lung, and Blood Institute (U01HL089897 and U01HL089856). The

project is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, Novartis, and Sepracor.

References

- Rabe KF, Hurd S, Anzueto A, et al, and the Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007; **176**: 532–55.
- Weatherall M, Marsh S, Shirtcliffe P, Williams M, Travers J, Beasley R. Quality of life measured by the St George's Respiratory Questionnaire and spirometry. *Eur Respir J* 2009; **33**: 1025–30.
- Hurst JR, Vestbo J, Anzueto A, et al, and the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) Investigators. Susceptibility to exacerbation in chronic obstructive pulmonary disease. *N Engl J Med* 2010; **363**: 1128–38.
- Agusti A, Calverley PM, Celli B, et al, and the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) investigators. Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respir Res* 2010; **11**: 122.
- Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. Global Initiative for Chronic Obstructive Lung Disease; 2011. <http://www.goldcopd.org/guidelines-global-strategy-for-diagnosis-management.html> (accessed April 4, 2012).
- Regan EA, Hokanson JE, Murphy JR, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD* 2010; **7**: 32–43.
- Brooks S. Surveillance for respiratory hazards. *ATS News* 1982; **8**: 12–16.
- Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Respir Med* 1991; **85** (suppl B): 25–31.
- Fletcher CM, Elmes PC, Fairbairn AS, Wood CH. The significance of respiratory symptoms and the diagnosis of chronic bronchitis in a working population. *BMJ* 1959; **2**: 257–66.
- Jones PW, Brusselle G, Dal Negro RW, et al. Properties of the COPD assessment test in a cross-sectional European study. *Eur Respir J* 2011; **38**: 29–35.
- Jones PW, Tabberer M, Chen WH. Creating scenarios of the impact of COPD and their relationship to COPD Assessment Test (CAT™) scores. *BMC Pulm Med* 2011; **11**: 42.
- Ferris BG. Epidemiology standardization project (American Thoracic Society). *Am Rev Respir Dis* 1978; **118**: 1–120.
- American Thoracic Society. 1994 Update. *Am J Respir Crit Care Med* 1995; **152**: 1107–36.
- Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med* 2004; **350**: 1005–12.
- Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *Am Rev Respir Dis* 1992; **145**: 1321–27.
- Mahler DA, Waterman LA, Ward J, McCusker C, ZuWallack R, Baird JC. Validity and responsiveness of the self-administered computerized versions of the baseline and transition dyspnea indexes. *Chest* 2007; **132**: 1283–90.
- Mahler DA, Ward J, Waterman LA, McCusker C, Zuwallack R, Baird JC. Patient-reported dyspnea in COPD reliability and association with stage of disease. *Chest* 2009; **136**: 1473–79.
- Westwood M, Bourbeau J, Jones PW, Cerulli A, Capkun-Niggli G, Worthy G. Relationship between FEV1 change and patient-reported outcomes in randomised trials of inhaled bronchodilators for stable COPD: a systematic review. *Respir Res* 2011; **12**: 40.
- Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidy N. Development and first validation of the COPD Assessment Test. *Eur Respir J* 2009; **34**: 648–54.
- Calverley PM, Anderson JA, Celli B, et al, and the TORCH investigators. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. *N Engl J Med* 2007; **356**: 775–89.
- Tashkin DP, Celli B, Senn S, et al, and the UPLIFT Study Investigators. A 4-year trial of tiotropium in chronic obstructive pulmonary disease. *N Engl J Med* 2008; **359**: 1543–54.
- Dodd JW, Marns PL, Clark AL, et al. The COPD Assessment Test (CAT): short- and medium-term response to pulmonary rehabilitation. *COPD* 2012; published online April 12. DOI:10.3109/15412555.2012.671869.
- Jones PW, Harding G, Wiklund I, et al. Tests of the responsiveness of the Chronic Obstructive Pulmonary Disease (COPD) Assessment Test TM (CAT) following acute exacerbation and pulmonary rehabilitation. *Chest* 2012; **142**: 134–40.
- Mackay AJ, Donaldson GC, Patel AR, Jones PW, Hurst JR, Wedzicha JA. Usefulness of the Chronic Obstructive Pulmonary Disease Assessment Test to evaluate severity of COPD exacerbations. *Am J Respir Crit Care Med* 2012; **185**: 1218–24.