Mortality prediction in chronic obstructive pulmonary disease comparing the GOLD 2015 and GOLD 2019 staging: a pooled analysis of individual patient data

Elena García Castillo[1,41], Tamara Alonso Pérez[1,41], Julio Ancochea [1,17], Maria Teresa Pastor Sanz[1], Pere Almagro [2], Pablo Martínez-Camblor[3], Marc Miravitlles [4], Mónica Rodríguez-Carballeira[2], Annie Navarro[5], Bernd Lamprecht[6], Ana S. Ramírez-García Luna [7], Bernhard Kaiser[8], Inmaculada Alfageme[9], Ciro Casanova[11], Cristóbal Esteban[11], Juan J. Soler-Cataluña[12], Juan P. de-Torres[13,14], Bartolomé R. Celli [15], Jose M. Marín [16,17], Gerben ter Riet[18,19], Patricia Sobradillo[20], Peter Lange[21], Judith Garcia-Aymerich [22,23,24], Josep M. Anto[22,23,24], Alice M. Turner[25], MeiLan K. Han[26], Arnulf Langhammer[27,26,28], Sigrid Anna Aalberg Vikjord [27], Alice Sternberg[29], Linda Leivseth[30], Per Bakke[31], Ane Johannessen[32], Toru Oga[33], Borja G. Cosío [34], Andrés Echazarreta [35], Nicolás Roche[36], Pierre-Régis Burgel [36], Don D. Sin[37,38], Milo A. Puhan[39], Jose Luis López-Campos [17,40], Laura Carrasco[40], Joan B. Soriano [1,17], for the 3CIA collaboration

ABSTRACT   In 2019, The Global Initiative for Chronic Obstructive Lung Disease (GOLD) modified the grading system for patients with COPD, creating 16 subgroups (1A–4D). As part of the COPD Cohorts Collaborative International Assessment (3CIA) initiative, we aim to compare the mortality prediction of the 2015 and 2019 COPD GOLD staging systems.

We studied 17 139 COPD patients from the 3CIA study, selecting those with complete data. Patients were classified by the 2015 and 2019 GOLD ABCD systems, and we compared the predictive ability for 5-year mortality of both classifications.

In total, 17 139 patients with COPD were enrolled in 22 cohorts from 11 countries between 2003 and 2017; 8823 of them had complete data and were analysed. Mean±SD age was 63.9±9.8 years and 62.9% were male. GOLD 2019 classified the patients in milder degrees of COPD. For both classifications, group D had higher mortality. 5-year mortality did not differ between groups B and C in GOLD 2015; in GOLD 2019, mortality was greater for group B than C. Patients classified as group A and B had better sensitivity and positive predictive value with the GOLD 2019 classification than GOLD 2015. GOLD 2015 had better sensitivity for group C and D than GOLD 2019. The area under the curve values for 5-year mortality were only 0.67 (95% CI 0.66–0.68) for GOLD 2015 and 0.65 (95% CI 0.63–0.66) for GOLD 2019.

The new GOLD 2019 classification does not predict mortality better than the previous GOLD 2015 system.

@ERSpublications
GOLD 2019 staging system created 16 subgroups. GOLD 2015 and GOLD 2019 are not strong predictors of mortality, and do not have sufficient discriminatory power to be used as a tool for risk classification of mortality in patients with COPD. https://bit.ly/3idBuaN

Introduction

COPD is a common cause of morbidity and mortality in the world. COPD affects ~328 million people worldwide, and COPD-related deaths amount to 4 million every year [1]. Assessment of disease severity is essential to predict prognosis and to standardise treatment regimes. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) document is the most widely used treatment guide for the staging and management of COPD. The GOLD grading system for COPD has significantly evolved since first publication in 2001 to the current version in 2019. Initially, in the GOLD 2007 classification, only post-bronchodilator airflow limitation based on spirometry forced expiratory volume in 1 s (FEV_1) was used to grade the severity of COPD [2]. Later on, some criticism arose on this grading score because it relied only upon FEV_1, which is not a good predictor of dyspnoea, quality of life or exercise tolerance. Further, other important variables to evaluate prognosis in COPD, such as sub-phenotypes, exacerbations, dyspnoea severity or comorbidities have been proposed, among others. Therefore, GOLD 2011 proposed a classification system of four groups, ABCD, combining FEV_1 and two clinical parameters: history of exacerbations and respiratory symptoms measured by the modified Medical Research Council (mMRC) dyspnoea score, or the COPD Assessment Test score (CAT) [3]. The 2011 ABCD classification was considered an improvement in the management of patients with COPD, providing an opportunity to further guide the individualised care of these patients. GOLD 2011 predicted future exacerbations better than GOLD 2007, but there was no difference in mortality predictions or respiratory outcomes [4–7]. In 2015 an updated report was published with the same measurement parameters (FEV_1, dyspnoea and exacerbations) than the 2011 classification [8].

The latest GOLD update, published in 2019, uses a composite of spirometry, symptoms and exacerbations, but importantly separating the spirometric 1–4 staging from the ABCD groups [9]. This separation is relevant because it is known there are differences in the rate of exacerbations for the most severe COPD patients, depending on whether the risk is based on pulmonary function tests, on the history of exacerbations or both [10]. All these classifications were initially designed not to assess prognosis, but to aid

Affiliations: [1]Pneumology Dept, Hospital Universitario de la Princesa, Instituto de Investigación Hospital Universitario de la Princesa (IISP), Universidad Autónoma de Madrid, Madrid, Spain. [2]Internal Medicine Department, Mútua Terrassa University Hospital, Barcelona, Spain. [3]Geisel School of Medicine at Dartmouth, Hanover, NH, USA. [4]Pneumology Dept, Hospital University Vall d´Hebron, CIBER de Enfermedades Respiratorias (CIBERES), Barcelona, Spain. [5]Pneumology Service, Hospital Universitari Mútua Terrassa, Barcelona, Spain. [6]Dept of Pulmonary Medicine, Kepler-University-Hospital, Faculty of Medicine, Johannes-Kepler-University Linz, Linz, Austria. [7]Facultad de Medicina, Universidad Autónoma de San Luis Potosí, San Luis Potosí, Mexico. [8]Dept of Pulmonary Medicine, Paracelsus Medical University Hospital, Salzburg, Austria. [9]Departamento de Medicina, Universidad de Sevilla, HU Virgen de Valme, Seville, Spain. [10]Pulmonology Department, Hospital Galdakao-Usansolo, Galdakao, Spain. [11]Pulmonary Department, Research Unit, Hospital Universitario Nuestra Señora de La Candelaria, Universidad de La Laguna, Tenerife, Spain. [12]Servicio de Neumología, Hospital Arnau de Vilanova, Valencia, Spain. [13]Clinica Universidad de Navarra, Pamplona, Spain. [14]Respirology and Sleep Medicine Division, Queen´s University, Kingston, Canada. [15]Pulmonary and Critical Care Medicine, Harvard University, Brigham and Women´s Hospital, Boston, MA, USA. [16]Hospital Universitario Miguel Servet, Zaragoza, Spain. [17]Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, Madrid, Spain. [18]Urban Vitality – Centre of Expertise, Faculty of Health, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands. [19]Dept of Cardiology, Amsterdam UMC, location Academic Medical Center, Amsterdam, The Netherlands. [20]University Hospital of Cruces in Barakaldo, Barakaldo, Spain. [21]Section of Social Medicine, Dept of Public Health, Copenhagen University, Copenhagen City Heart Study, Frederiksberg Hospital, Copenhagen, Denmark. [22]ISGlobal, Barcelona, Spain. [23]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [24]CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain. [25]Institute of Applied Health Research, University of Birmingham, Edgbaston, UK. [26]Internal Medicine, Division of Pulmonary and Critical Care Medicine, University of Michigan, Ann Arbor, MI, USA. [27]Dept of Public Health and Nursing, NTNU (Norwegian University of Science and Technology), Trondheim, Norway. [28]Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway. [29]Dept of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [30]Centre for Clinical Documentation and Evaluation, Northern Norway Regional Health Authority, Tromso, Norway. [31]Dept of Clinical Science, University of Bergen, Bergen, Norway. [32]Dept of Global Public Health and Primary Care, University of Bergen, Bergen, Norway. [33]Dept of Respiratory Care and Sleep Control Medicine, Kyoto University, Kyoto, Japan. [34]Hospital Universitario Son Espases-IdISPa, Mallorca, Spain. [35]Servicio de Neumología, Hospital San Juan de Dios de La Plata, Buenos Aires, Argentina. [36]Respiratory Medicine, Cochin Hospital, APHP Centre–University of Paris, Cochin Institute (INSERM UMR1016), Paris, France. [37]UBC Centre for Heart Lung Innovation, Vancouver, BC, Canada. [38]Division of Respiratory Medicine, Dept of Medicine, University of British Columbia, Vancouver, BC, Canada. [39]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland. [40]Unidad Médico Quirúrgica de Enfermedades Respiratorias, Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/Universidad de Sevilla, Seville, Spain. [41]These authors contributed equally.

Correspondence: Joan B. Soriano, Hospital Universitario de la Princesa, Diego de León 62, Neumología 6ª planta, 28006-Madrid, Spain. E-mail: jbsoriano2@gmail.com

clinicians in creating optimal treatment regimens for patients. Thus, the prognostic ability of GOLD 2019 compared to previous classifications is largely unknown, with only a few published studies [11, 12]. To address this issue, we used pooled data from 17 139 patients of 22 COPD cohorts and 11 countries and compared the prognostic capacity of the 2019 *versus* 2015 GOLD staging classifications to predict mortality.

Methods

Study population

In this international study, we assessed 17 139 patients from the COPD Cohorts Collaborative International Assessment (3CIA) initiative. All were prospective cohorts that recruited patients within the period 1999 to 2017, except for one which was a population-based cohort. All patients had a definition of COPD characterised by spirometry, that is post-bronchodilator FEV_1 to forced vital capacity (FVC) ratio <0.7 and a clinical diagnosis of COPD. Spirometry was performed using the standards provided by the American Thoracic Society and European Respiratory Society [13]. The primary investigators of each of the participating 3CIA cohorts provided individual patient's data for pooled analysis. We obtained a minimum individual dataset including the vital status (up to death, right truncation or 2017), age, sex, smoking status, pre-bronchodilator and post-bronchodilator FEV_1 and FVC, and dyspnoea measured with the modified mMRC, among others [14]. Only in a number of 3CIA cohorts were data of number of exacerbations in the previous year available. For the current study, we selected exclusively those cohorts in which the number of exacerbations in the previous year were available in the database, since this variable is required to calculate the GOLD 2015 and 2019 grading systems. Fifteen out of a total of 22 cohorts contained data on history of exacerbations, so that the final number of patients available to be classified according to GOLD 2015 and GOLD 2019 was 8823. Symptoms were assessed using the mMRC dyspnoea scale. To determine the risk descriptor in the 2015 grouping system, we used exacerbations history and GOLD spirometry stages. The combination of symptoms (mMRC) and the worse risk descriptor (spirometry or exacerbation history) were used to classify patients by the GOLD 2015 system. Participants were classified using the GOLD 2019 system into four grades (1–4) based on post-bronchodilator FEV_1 percentage of prediction as stage 1 ($FEV_1 \geqslant 80$), stage 2 (FEV_1 79–50), stage 3 (FEV_1 30–49) and stage 4 (FEV_1 <30). Groups ABCD were defined by self-reported severity of dyspnoea (mMRC) and number of exacerbations in the previous year.

All participants provided informed written consent, and each study was conducted with the formal approval of the local ethics institutional committees following the principles of the Declaration of Helsinki.

Outcomes

The primary outcome was the prediction ability of all-cause mortality in the individuals by the two GOLD systems.

Statistical analysis

The 3CIA database manager quality-controlled all data centrally and created a clean database with a data dictionary. All implausible or missing variables were queried with the original study investigators, and data were removed from the central database if errors could not be corrected. Because the cohorts had different follow-up times, patients were right-censored at 5 years of follow-up.

Descriptive statistics used mean and standard deviation for continuous variables and the number of cases and percentages for categorical variables. Comparisons between groups were performed with the Chi-squared test for categorical variables and the t-test for continuous variables.

We estimated 5-year all-cause mortality, according to GOLD 2015 and 2019 staging systems, using Kaplan–Meier survival statistics. Statistical comparisons were performed using the log-rank test receiver operation characteristic (ROC) curve analyses and area under the curve (AUC) and the 95% confidence intervals of the AUC were calculated to measure the predictive accuracy for mortality. Also, we compared the prediction ability of mortality on both classifications using sensitivity, positive predictive value and the Youden's index with Epidat 3.1 programme.

Results

We pooled data from 22 COPD cohorts with a total of 17 139 patients, finally including 8823 patients from 15 cohorts that had all complete variables to be classified as GOLD 2015 and GOLD 2019. A comparison of baseline characteristics of included and not included patients is presented in table 1. There were statistically significant differences in many variables given the large size, but most should be considered not clinically relevant (table 1).

The 8823 included patients were 62.9% male, with a mean±SD age of 63.9 ±9.8 years, body mass index 27.0±5.8 kg·m^{-2} and mMRC dyspnoea score of 1.8±1.4. Post-bronchodilator FEV_1 was 54.8%±22.3 of the

TABLE 1 Comparison of demographic and clinical characteristics in COPD Cohorts Collaborative International Assessment (3CIA) COPD patients included/excluded in this analysis

	Excluded	Included	p-value
Subjects n	8316	8823	
Age years	64.2±10.7	63.9±9.8	0.08
Male sex	6232 (74.9%)	5552 (62.9%)	<0.001
BMI kg·m^{-2}	26.5±4.9	27.0±5.8	<0.001
Modified MRC dyspnoea scale	1.5±1.3	1.8±1.4	<0.001
0	1647 (23.4%)	1957 (22.2%)	<0.001
1	2261(32.1%)	1886 (21.4%)	
2	1641 (23.3%)	1772 (20.1%)	
3	688 (9.8%)	1951 (22.1%)	
4	805 (11.4%)	1257 (14.3%)	
Six-min walk distance m	415.4±108.8	376.9±129.1	<0.001
FEV$_1$ post BD mL	1.7±0.7	1.6±0.8	<0.001
FEV$_1$ post BD %	60.8±22.1	54.8±22.3	<0.001
Smoker			<0.001
Former	3989 (49.1%)	5392 (61.4%)	
Current	3589 (44.2%)	3174 (36.1%)	
Never	542 (6.7%)	222 (2.5%)	
Pack-years	46.4±28.8	42.1±28.3	<0.001
Cough	1103 (54.2%)	342 (43.9%)	<0.001
Sputum	1159 (42.1%)	341 (43.9%)	0.353
Diabetes	354 (6.7%)	303 (16.6%)	<0.001
Cardiac disease	1072 (30.8%)	467 (25.9%)	<0.001
Chronic bronchitis	166 (38.7%)	787 (69.5%)	<0.001
Hypertension	454 (40.4%)	826 (44.8%)	0.028
Asthma	1243 (26.1%)	209 (10.7%)	<0.001
Spirometry staging			<0.001
1	1567 (19%)	1153 (13.1%)	
2	3892 (47.1%)	3711 (42.1%)	
3	2126 (25.8%)	2654 (30.1%)	
4	671 (8.1%)	1301 (14.8%)	
Long-term oxygen therapy	119 (1.4%)	430 (4.8%)	0.259

Data are presented as n (%), mean±SD or median (interquartile range), unless otherwise stated. BMI: Body mass index; MRC: Medical Research Council; FEV$_1$: forced expiratory volume in 1 s; BD: bronchodilator.

predictive value, and 6-min walk distance was 376.9±129.1 m. Based on spirometry staging, 1153 (13%) had mild (stage 1), 3711 (42.1%) had moderate (stage 2), 2654 (30.1%) had severe (stage 3) and 1301 (14.8%) had very severe (stage 4) disease.
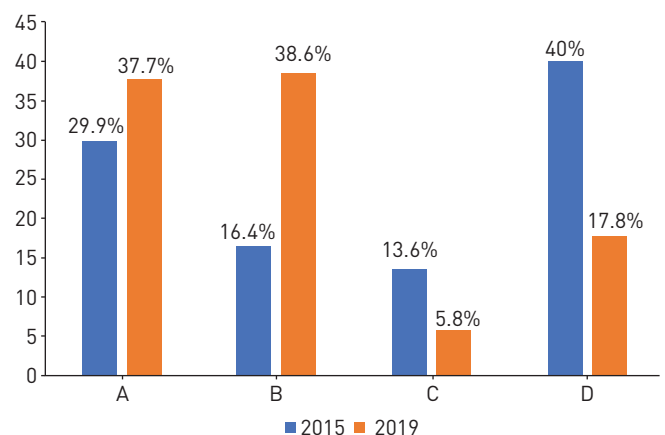

FIGURE 1 Distribution of participants by classification in Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2015 and GOLD 2019.

TABLE 2 Mortality risk among COPD patients according to Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2019 and GOLD 2015 classifications

	GOLD 2015 5-year mortality %	GOLD 2019 5-year mortality %
Group A	5.8	7.5
Group B	13.8	23.2
Group C	14.1	14.8
Group D	30.8	32.8

The distribution of these 8823 patients according to GOLD 2015 and GOLD 2019 is presented in figure 1. With GOLD 2019 there is a shift towards the less severe staging of disease (absolute increase in stage A and B of 7.8% and 22.2% respectively; and absolute decrease in stage C and D of 7.8% and 22.2% respectively).

The overall 5-year mortality rate was 18.3%. The all-cause 5-year mortality rates according to both classifications are shown in table 2. Figure 2 shows Kaplan–Meier curves for 5-year mortality according to GOLD 2015 (figure 2a) and GOLD 2019 (figure 2b). All-cause mortality at 5 years in the 2015 GOLD classification was higher in grade D, followed by grades B, C (with similar mortality), and finally grade A, log-rank test p<0.001 (table 2 and figure 2a). Grade D diverged from the beginning of follow-up, while grades B and C diverged after 1 year of follow-up. In GOLD 2019, the four Kaplan–Meier curves diverge during the first year, but interestingly, grade B had higher mortality than grade C, so mortality was higher in groups B and D (more symptoms) than in groups A and C (fewer symptoms; figure 2b). The degree of obstruction measured by FEV_1% further subclassified patients into 16 subgroups with different mortality rates, increasing from 1A to 4D in the GOLD 2019 grading system (table 3). Figure 3 shows Kaplan–Meier curves for each of the spirometry strata. The higher mortality of group B over group C persisted in each of the strata with the exception of spirometry strata 1, with a higher mortality in group C. Similarly, to GOLD 2015, in GOLD 2019, grades A and D had the lowest and highest mortality, respectively, with very similar absolute numbers (table 2, figure 2).

The primary outcome, the prediction capacity as measured by the AUC of ROC curve for mortality up to 5 years, was intermediate (<0.70) for both systems (figure 4). GOLD 2015 exhibited slightly better discrimination in predicting mortality (AUC 0.67, 95% CI 0.66–0.68) than the GOLD 2019 classification (AUC 0.64, 95% CI 0.63–0.66).

Patients at risk

a) GOLD 2015

	0	12	24	36	48	60
A	2388	2285	2129	1914	1420	937
B	1259	1184	1045	876	608	365
C	1129	1070	968	875	693	563
D	3229	2962	2609	2128	1525	977

b) GOLD 2019

	0	12	24	36	48	60
A	3039	2909	2687	2419	1806	1238
B	3053	2854	2539	2100	1453	873
C	478	446	410	370	307	262
D	1435	1292	1115	904	680	469

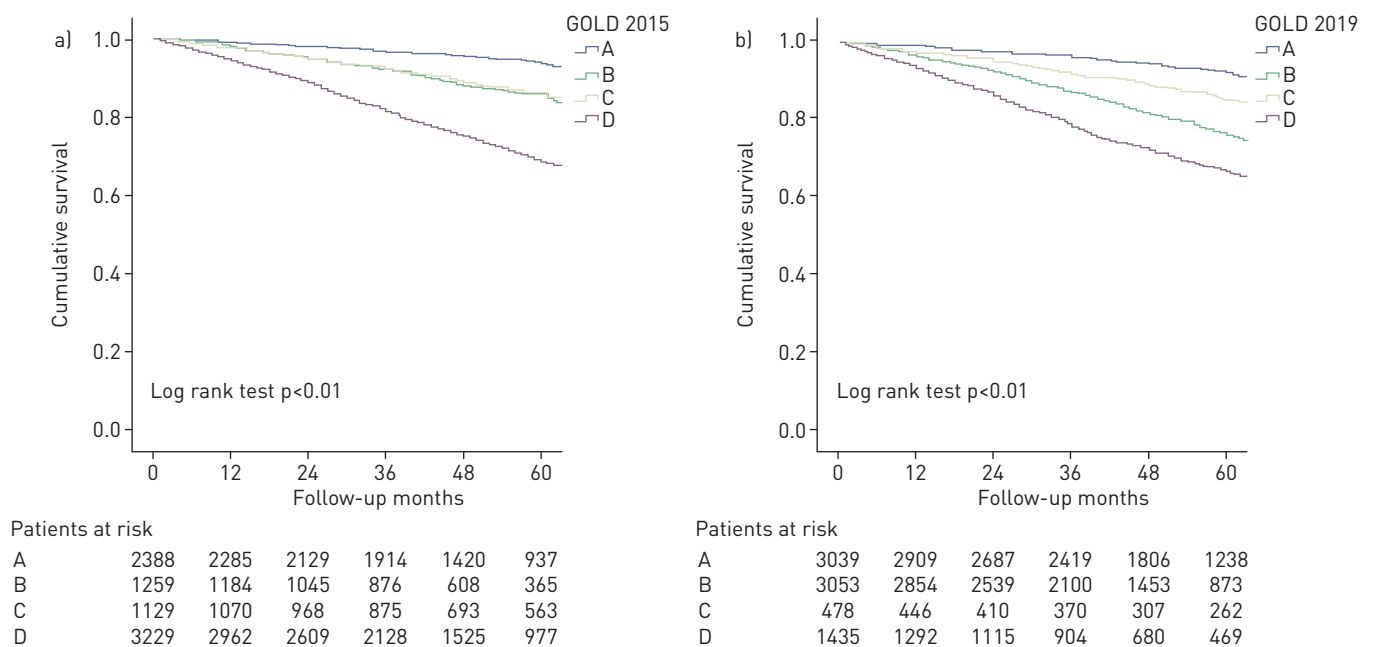FIGURE 2 Kaplan–Meier survival curves by a) Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2015 and b) GOLD 2019.

GOLD 2019	A	B	C	D
Spirometry I	3.4	9.8	16.2	5.5
Spirometry II	7	14.9	8.3	22.1
Spirometry III	13	23.7	17.7	32.6
Spirometry IV	16.7	39.1	29.5	46

TABLE 3 Five-year all-cause mortality (%) among spirometry strata within Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2019 ABCD classification
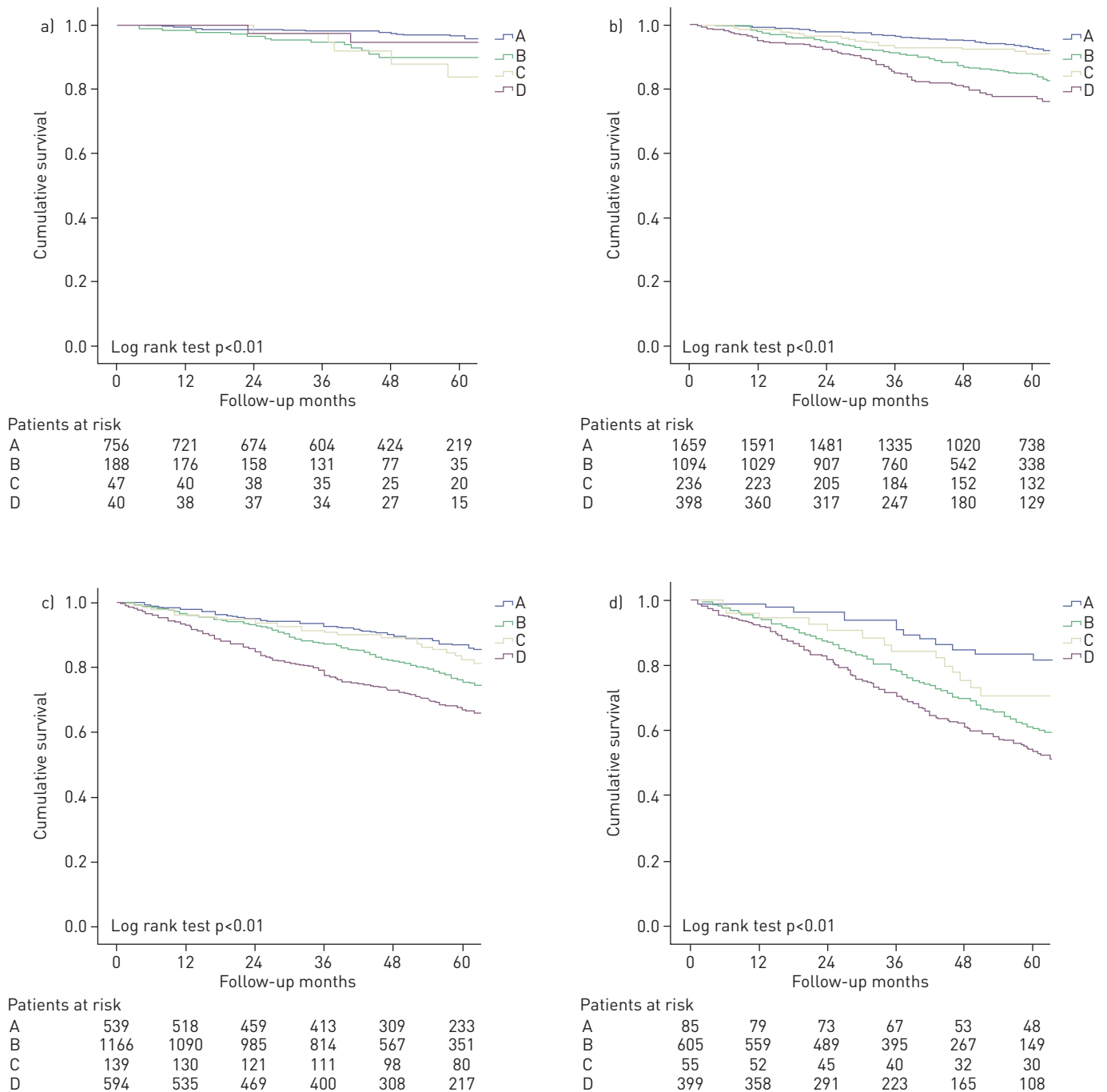

FIGURE 3 Kaplan–Meier survival curves by Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2019 and spirometry subgroups. a) Spirometry subgroup I; b) spirometry subgroup II; c) spirometry subgroup III; d) spirometry subgroup IV.
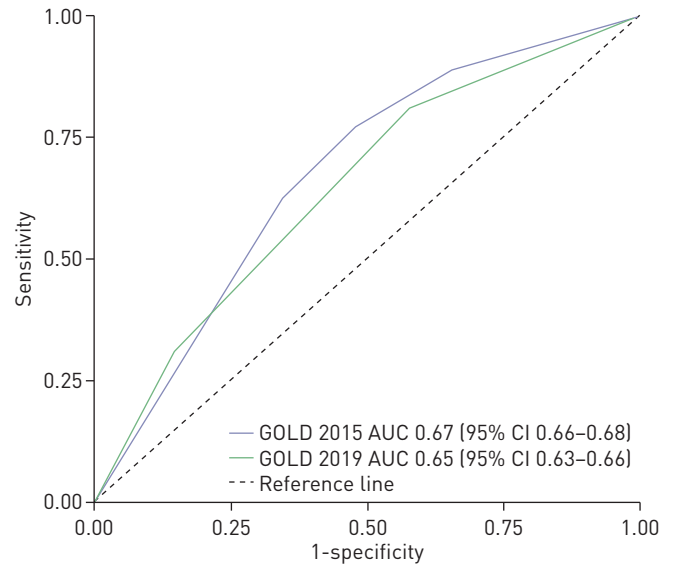
FIGURE 4 Receiver operating curves for all-cause mortality at 5 years follow-up.

Regarding sensitivity parameters, both classifications had quite shallow values. GOLD 2019 had a higher sensitivity for predicting mortality on A and B groups *versus* the 2015 classification (19.1 *versus* 11.1 and 43.6 *versus* 11.6). On the other hand, GOLD 2015 had a higher sensitivity on groups C and D *versus* the 2019 classification (14.6 *versus* 6.6 and 62.5 *versus* 30.6) based on overlapping 95% confidence intervals (table 4). The positive predictive values were also higher in GOLD 2019 group A and B *versus* the same groups in GOLD 2015 (9.5 *versus* 6.9 and 21.2 *versus* 13.3); but not different (overlapping confidence intervals) in groups C and D. The Youden indices were quite low for both classifications, even with negative values, showing that it is not an optimal classification system to assess mortality.

Discussion

Our study evaluates mortality according to the last two GOLD classifications, and it is one of the most extensive to date. The most important finding is that the new GOLD 2019 classification (based on symptoms or exacerbations along with stages spirometry analysed altogether) did not predict 5-year mortality better than the GOLD 2015 classification (based on spirometry, history of exacerbation and symptoms). How to define and stage COPD exacerbations is a matter of intense, long debate [15]. Most 3CIA individual cohorts used subsequent iterations of GOLD-accepted definitions in their protocols, mostly based on RODRIGUEZ-ROISIN *et al.*'s [16] seminal paper, including mild (symptom-only) exacerbations. However, when pooling for 3CIA, we only focused on those COPD exacerbations that required health services use, emergency room admission or death. In the GOLD 2019 classification, subgroups B and D had the worst mortality, highlighting that the higher burden of symptoms conveys a higher mortality independently of spirometry (table 2 and figure 2b). Finally, we show an important shift in the proportions of patients between the 2015 and 2019 ABCD grades, with milder severity in GOLD 2019.

TABLE 4 Accuracy for predicting mortality of the ABCD groups classifications by Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2015 and GOLD 2019 schemes

Classification	Sensitivity (95% CI)	PPV (95% CI)	Youden index (95% CI)
GOLD 2015 group A	11.1 (9.6–12.7)	6.9 (5.9–7.9)	−0.23 (−0.25−−0.22)
GOLD 2015 group B	11.6 (10–13.2)	13.3 (11.5–15.1)	−0.06 (−0.08−−0.04)
GOLD 2015 group C	14.6 (12.8–16.4)	20.2 (17.9–22.6)	+0.01 (−0.01−+0.03)
GOLD 2015 group D	62.5 (60.1–64.9)	29.5 (28–31.1)	+0.28 (+0.25−+0.31)
GOLD 2019 group A	19.1 (17.2–21.1)	9.5 (8.4–10.5)	−0.23 (−0.25−−0.21)
GOLD 2019 group B	43.6 (41.1–46.1)	21.2 (19.8–22.6)	+0.06 (+0.03−+0.09)
GOLD 2019 group C	6.6 (5.3–7.8)	21.7 (18–25.5)	+0.01 (0.0−+0.02)
GOLD 2019 group D	30.6 (28.3–32.8)	32.8 (30.4–35.2)	+0.16 (+0.14−+0.18)

PPV: positive predictive value.

GOLD 2015 and GOLD 2019 classifications had a low discriminatory power as per the AUCs, ranging from 0.67 to 0.65, and similar to other studies (by consensus, AUCs below 0.70 are considered low or weak) [11]. Sensitivity and positive predictive values indicate that the general performance of the two models is similar and very low. Also, the Youden's indices are very low with negative values that have no meaningful interpretation in practice. These findings support the results of other studies, suggesting that GOLD classification is not a good predictor for mortality [11, 12]. There may be various reasons for these poor results. The main reason is that these classifications were conceived to guide treatment, so it is not surprising that their capacity for predicting mortality is low. There are clinical phenotypes such as the asthma-COPD, the frequent exacerbator with emphysema or chronic bronchitis comorbidities and different indexes, that are significant predictors of mortality and are not entirely included in the GOLD stages [17–21]. In our study, sensitivity and positive predictive value are higher in GOLD 2019 groups A and B. On the contrary, sensitivity is higher in GOLD 2015 in groups C and D. These results suggest that GOLD 2019 predicts slightly better mortality in low-risk groups (A and B) and GOLD 2015 in high-risk groups (C and D), requiring more studies to corroborate it.

The discriminatory power of GOLD 2019 was lower than GOLD 2015 as shown in AUC values. The partition of FEV_1 as a direct classifier in GOLD 2019 reduced ability to discriminate survival, highlighting the need to consider the severity of airflow obstruction in assessing mortality risk. When using GOLD 2019 with a composite of spirometry, exacerbations and symptoms (16 subgroups 1A to 4D classification), we found an increase in all-cause mortality between GOLD 2019 stage 1 and GOLD 2019 stage 4 across grades A, B and D, highlighting the persisting importance of FEV_1 as a predictor of mortality (table 3 and figure 3). In group C, mortality was higher in spirometry stage 1 than in stage 2, probably due to a significant difference of the proportion in patients between the two stages.

Our study confirms that patients classified as GOLD A had the best survival, and patients with GOLD D had the higher mortality in both classifications [22–24]. Mortality of groups B and C in GOLD 2015 lay in between A and D groups and often overlapped. In GOLD 2019 mortality was significantly higher in group B patients than in group C. This finding is similar to other previous reports published [6, 23, 25], showing that group B is an intermediate- to high-risk group associated with more exacerbations and likely with other comorbidities that may cause dyspnoea (such as heart failure). Furthermore, we show that the burden of symptoms (represented by groups B and D) have a prognostic value additive but independent to spirometry.

The current study demonstrates that the use of the GOLD 2019 classification scheme shifted patients with COPD to groups of milder severity compared with GOLD 2015. This happened in keeping with previous reports, but in a smaller proportion of patients (30% of patients reassigned to group A or B compared to 53% in LEE et al. [25] study or 66% in TAN et al. [26] study). The further distribution of spirometric parameters from two categories in GOLD 2015 (FEV_1 less or higher than 50%) to four categories in GOLD 2019 was one of the possible reasons for the patients' shift from C and D in GOLD 2015 to A and B groups in GOLD 2019. This phenomenon is opposite to the one observed with the use of the revised GOLD 2011 classification, which shifted the patients from the GOLD 2007 classification towards more advanced stages of the disease (D group increased almost three times) [7]. Remarkably, in both classifications, group C was consistently the smallest group (figure 1), as seen in other reports [23]. The implication of progressively milder disease classifications for the treatment choices clinicians make in real life practice guided by GOLD is not yet clear given the recent nature of the latest GOLD iteration but will be important to monitor.

The strengths of our study include a large sample size, the study design (a pooled analysis of individual patient data from several cohorts) and the different degrees of severity of patients from different cohorts, maximising its high external validity. Prospective data collection of spirometry with a post-bronchodilator test, dyspnoea by mMRC scale, history of exacerbation and mortality enabled direct classification of patients by the 2015 and 2019 GOLD staging schemes. We also have a significant representation of women, which other COPD studies might not have achieved [28].

Our study has several limitations. First, although we started with 17 139 patients with COPD, a considerable number were excluded because of missing information on key variables, mainly regarding the history of exacerbations. These missing data are unlikely to affect the validity of our results, as we can see in our analysis comparing included and non-included patients. Second, the mortality analysis used all-cause death, and we have no data regarding specific causes of death (this data was not collected consistently in all cohorts). Third, symptoms were only evaluated using the mMRC dyspnoea score, but not with the COPD Assessment Test, or other instruments [29]; however, that is in line with other reported cohorts [30, 31]. Fourth, most patients came from hospital-based cohorts, so we likely have more patients with moderate to severe disease and less patients with mild and moderate disease compared to an

outpatient setting, or a primary care population. Indeed, the 22 cohorts from 11 countries in 3CIA within our initiative are only a sample representing the estimated 300 million COPD patients worldwide [32]. Finally, the recruitment timeframe was long (10 years), so evolving treatment recommendations might influence results.

In conclusion, this study of COPD cohorts, including 8823 patients, showed that neither GOLD 2015 nor GOLD 2019 are strong predictors of mortality. GOLD 2019 predicted mortality better than GOLD 2015 in groups A and B but worse in groups C and D. However, none of the GOLD classifications has sufficient discriminatory power to be used as a tool for risk classification of mortality in patients with COPD. Ours should be considered a constructive exercise and a critical appraisal of the last two GOLD iterations defining COPD. Within 3CIA, we have already suggested several proposals for future COPD staging and grading classifications, by applying more evidence-based thresholds of evidence-based variables [7, 21, 33–35].

Conflict of interest: None declared.

References

1   GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir Med* 2017; 5: 691–706.
2   Rabe KF, Hurd S, Anzueto A, *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007; 176: 532–555.
3   Vestbo J, Hurd SS, Agustí AG, *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2013; 187: 347–365.
4   Han MK, Muellerova H, Curran-Everett D, *et al.* GOLD 2011 disease severity classification in COPDGene: a prospective cohort study. *Lancet Respir Med* 2013; 1: 43–50.
5   Soriano JB, Alfageme I, Almagro P, *et al.* Distribution and prognostic validity of the new Global Initiative for Chronic Obstructive Lung Disease grading classification. *Chest* 2013; 143: 694–702.
6   Lange P, Marrot 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.
7   Soriano JB, Lamprecht B, Ramírez AS, *et al.* Mortality prediction in chronic obstructive pulmonary disease comparing the GOLD 2007 and 2011 staging systems: a pooled analysis of individual patient data. *Lancet Respir Med* 2015; 3: 443–450.
8   Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the diagnosis, management and prevention of COPD 2015. www.goldcopd.org/ Date last accessed: 21 June 2019, Date last updated: 1 October 2020.
9   Singh D, Agusti A, Anzueto A, *et al.* Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease: the GOLD science committee report 2019. *Eur Respir J* 2019; 53: 1900164.
10   Goossens LM, Leimer I, Metzdorf N, *et al.* Does the 2013 GOLD classification improve the ability to predict lung function decline, exacerbations and mortality: a post-hoc analysis of the 4-year UPLIFT trial. *BMC Pulm Med* 2014; 14: 163.
11   Gedebjerg A, Szépligeti SK, Wackerhausen LH, *et al.* Prediction of mortality in patients with chronic obstructive pulmonary disease with the new Global Initiative for Chronic Obstructive Lung Disease 2017 classification: a cohort study. *Lancet Respir Med* 2018; 6: 204–212.
12   Aramburu A, Arostegui I, Moraza J, *et al.* COPD classification models and mortality prediction capacity. *Int J Chron Obstruct Pulmon Dis* 2019; 14: 605–613.
13   Miller MR, Hankinson J, Brusasco V, *et al.* Standardisation of spirometry. *Eur Respir J* 2005; 26: 319–338.
14   Bestall JC, Paul EA, Garrod R, *et al.* Usefulness of the Medical Research Council (MRC) dyspnea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax* 1999; 54: 581–586.
15   Mathioudakis AG, Moberg M, Janner J, *et al.* Outcomes reported on the management of COPD exacerbations: a systematic survey of randomised controlled trials. *ERJ Open Res* 2019; 5: 00072-2019.
16   Rodriguez-Roisin R. Toward a consensus definition for COPD exacerbations. *Chest* 2000; 117: 5 Suppl. 2, 398S–401S.
17   Golpe R, Suárez-Valor M, Martín-Robles I, *et al.* Mortality in COPD patients according to clinical phenotypes. *Int J Chron Obstruct Pulmon Dis* 2018; 13: 1433–1439.
18   Marin JM, Alfageme I, Almagro P, *et al.* Multicomponent indices to predict survival in COPD: the COCOMICS study. *Eur Respir J* 2013; 42: 323–332.
19   Izquierdo JL, Miravitlles M, Esquinas C, *et al.* Characteristics of COPD patients managed in respiratory medicine departments in Spain, according to GOLD groups and GesEPOC clinical phenotypes. *Arch Bronconeumol* 2018; 54: 559–567.
20   Radovanovic D, Contoli M, Marco FD, *et al.* Clinical and functional characteristics of COPD patients across GOLD classifications: results of a multicenter observational study. *COPD* 2019; 16: 215–226.
21   Almagro P, Martínez-Camblor P, Miravitlles M, *et al.* External validation and recalculation of the CODEX Index in COPD patients. A 3CIAplus cohort study. *COPD* 2019; 16: 8–17.
22   Han MZ, Hsiue TR, Tsai SH, *et al.* Validation of the GOLD 2017 and new 16 subgroups (IA-4D) classifications in predicting exacerbation and mortality in COPD patients. *Int J Chron Obstruct Pulmon Dis* 2018; 13: 3425–3433.
23   Flynn RW, MacDonald TM, Chalmers JD, *et al.* The effect of changes to GOLD severity stage on long term morbidity and mortality in COPD. *Respir Res* 2018; 19: 249.

24   de Torres JP, Casanova C, Marin JM, *et al.* Prognostic evaluation of COPD patients: GOLD 2011 versus BODE and the COPD comorbidity index COTE. *Thorax* 2014; 69: 799–804.

25   Lee SJ, Yun SS, Ju S, *et al.* Validity of the GOLD 2017 classification in the prediction of mortality and respiratory hospitalization in patients with chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2019; 14: 911–919.

26   Tan WC, Bourbeau J, Aaron SD, *et al.* GOLD 2017 classification and lung function decline in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2018; 197: 670–673.

27   Criner RN, Labaki WS, Regan EA, *et al.* Mortality and exacerbations by Global Initiative for Chronic Obstructive Lung Disease Groups ABCD: 2011 versus 2017 in the COPDGene Cohort. *Chronic Obstr Pulm Dis* 2019; 6: 64–73.

28   Kobayashi S, Hanagama M, Ishida M, *et al.* Clinical characteristics and outcomes in Japanese patients with COPD according to the 2017 GOLD classification: the Ishinomaki COPD Network Registry. *Int J Chron Obstruct Pulmon Dis* 2018; 13: 3947–3955.

29   Nathan RA, Sorkness CA, Kosinski M, *et al.* Development of the Asthma Control Test: a survey for assessing asthma control. *J Allergy Clin Immunol* 2004; 113: 59–65.

30   Johannessen A, Nilsen RM, Storebo 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.

31   Agusti A, Edwards LD, Celli B, *et al.* Characteristics, stability and outcomes of the 2011 GOLD COPD groups in the ECLIPSE cohort. *Eur Respir J* 2013; 42: 636–646.

32   GBD Chronic Respiratory Disease Collaborators. Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Respir Med* 2020; 8: 585–596.

33   Guerra B, Haile SR, Lamprecht B, *et al.* Large-scale external validation and comparison of prognostic models: an application to chronic obstructive pulmonary disease. *BMC Med* 2018; 16: 33.

34   Haile SR, Guerra B, Soriano JB, *et al.* Multiple Score Comparison: a network meta-analysis approach to comparison and external validation of prognostic scores. *BMC Med Res Methodol* 2017; 17: 172.

35   Burgel PR, Paillasseur JL, Janssens W, *et al.* A simple algorithm for the identification of clinical COPD phenotypes. *Eur Respir J* 2017; 50: 1701034.