Mortality prediction in chronic obstructive pulmonary disease comparing the GOLD 2007 and 2011 staging systems: a pooled analysis of individual patient data

Joan B Soriano, Bernd Lamprecht, Ana S Ramírez, Pablo Martinez-Camblor, Bernhard Kaiser, Inmaculada Alfageme, Pere Almagro, Ciro Casanova, Cristobal Esteban, Juan J Soler-Cataluña, Juan P de-Torres, Marc Miravitlles, Bartolome R Celli, Jose M Marin, Milo A Puhan, Patricia Sobradillo, Peter Lange, Alice L Sternberg, Judith Garcia-Aymerich, Alice M Turner, MeiLan K Han, Arnulf Langhammer, Linda Leivseth, Per Bakke, Ane Johannessen, Nicolas Roche, Don D Sin

Summary

Background There is no universal consensus on the best staging system for chronic obstructive pulmonary disease (COPD). Although documents (eg, the Global Initiative for Chronic Obstructive Lung Disease [GOLD] 2007) have traditionally used forced expiratory volume in 1 s (FEV_1) for staging, clinical parameters have been added to some guidelines (eg, GOLD 2011) to improve patient management. As part of the COPD Cohorts Collaborative International Assessment (3CIA) initiative, we aimed to investigate how individual patients were categorised by GOLD 2007 and 2011, and compare the prognostic accuracy of the staging documents for mortality.

Methods We searched reports published from Jan 1, 2008, to Dec 31, 2014. Using data from cohorts that agreed to participate and had a minimum amount of information needed for GOLD 2007 and 2011, we did a patient-based pooled analysis of existing data. With use of raw data, we recalculated all participant assignments to GOLD 2007 I–IV classes, and GOLD 2011 A–D stages. We used survival analysis, C statistics, and non-parametric regression to model time-to-death data and compare GOLD 2007 and GOLD 2011 staging systems to predict mortality.

Findings We collected individual data for 15 632 patients from 22 COPD cohorts from seven countries, totalling 70 184 person-years. Mean age of the patients was 63·9 years (SD 10·1); 10 751 (69%) were men. Based on FEV_1 alone (GOLD 2007), 2424 (16%) patients had mild (I), 7142 (46%) moderate (II), 4346 (28%) severe (III), and 1670 (11%) very severe (IV) disease. We compared staging with the GOLD 2007 document with that of the new GOLD 2011 system in 14 660 patients: 5548 (38%) were grade A, 2733 (19%) were grade B, 1835 (13%) were grade C, and 4544 (31%) were grade D. GOLD 2011 shifted the overall COPD severity distribution to more severe categories. There were nearly three times more COPD patients in stage D than in former stage IV ($p<0·05$). The predictive capacity for survival up to 10 years was significant for both systems ($p<0·01$) but area under the curves were only 0·623 (GOLD 2007) and 0·634 (GOLD 2011), and GOLD 2007 and 2011 did not differ significantly. We identified the percent predicted FEV_1 thresholds of 85%, 55% and 35% as better to stage COPD severity for mortality, which are similar to the ones used previously.

Interpretation Neither GOLD COPD classification schemes have sufficient discriminatory power to be used clinically for risk classification at the individual level to predict total mortality for 3 years of follow-up and onwards. Increasing intensity of treatment of patients with COPD due to their GOLD 2011 reclassification is not known to improve health outcomes. Evidence-based thresholds should be searched when exploring the prognostic ability of current and new COPD multicomponent indices.

Funding None.

Lancet Respir Med 2015; 3: 443–50

Published Online May 18, 2015 http://dx.doi.org/10.1016/S2213-2600(15)00157-5

See Comment page 418

See Online for a podcast interview with Joan Soriano

Instituto de Investigación Hospital Universitario de la Princesa (IISP), Universidad Autónoma de Madrid, Cátedra UAM-Linde, Madrid, Spain (J B Soriano MD); **Department of Pulmonary Medicine, Kepler-University-Hospital, Linz, Austria** (B Lamprecht MD); **Faculty of Medicine, Johannes-Kepler-University, Linz, Austria** (B Lamprecht); **Facultad de Medicina UASLP, San Luis Potosí, Mexico** (A S Ramírez); **Universidad Autónoma de Chile, Chile** (P Martinez-Camblor PhD); **Hospital Universitario Central de Asturias (HUCA), Oviedo, Spain** (P Martinez-Camblor); **Department of Pulmonary Medicine, Paracelsus Medical University Hospital, Salzburg, Austria** (B Kaiser PhD); **Hospital Universitario Valme, Seville, Spain** (I Alfageme MD); **Internal Medicine, Hospital Universitari Mutua de Terrassa, Universitat de Barcelona, Barcelona, Spain** (P Almagro MD); **Hospital Nuestra Señora de la Candelaria, Tenerife, Spain** (C Casanova MD); **Hospital Galdakao-Usansolo, Galdakao, Bizkaia, Spain** (C Esteban MD); **Servicio de Neumología, Hospital Arnau de Vilanova, Valencia, Spain** (J J Soler-Cataluña MD); **Clinica Universidad de Navarra, Pamplona, Spain** (J P de-Torres MD); **Pneumology Department, Hospital Universitary Vall d'Hebron, CIBER de Enfermedades Respiratorias (CIBERES),**

Introduction

Chronic obstructive pulmonary disease (COPD) affects about 328 million people worldwide and accounts for 4 million deaths every year.[1] To address the growing global burden of COPD, efforts have been made to increase awareness and to standardise treatment models. The most influential effort has been the Global Initiative for Chronic Obstructive Lung Disease (GOLD), which was launched in 1997 as a joint effort between the US National Institutes of Health and WHO.[2] Traditionally, GOLD and other similar initiatives have classified patients' COPD severity through thresholds of forced expiratory volume in 1 sec (FEV_1). However, in 2011 the GOLD committee added clinical parameters (ie, a history of exacerbation and respiratory symptoms) to FEV_1 to improve clinical management of patients with COPD.[3] Although the GOLD staging system was designed to guide therapy, it is commonly used for prognostication.[4,5] The usefulness of both the 2007 and 2011 GOLD staging systems to predict mortality of patients with COPD is debated, with small, partial studies giving largely inconclusive results. To resolve this important issue, we obtained and pooled individual data for 15 632 patients from 22 COPD

Barcelona, Spain (M Miravitlles MD); Harvard University, Brigham and Women's Hospital, Pulmonary and Critical Care Medicine, Boston, MA, USA (B R Celli MD); Hospital Universitario Miguel Servet, Zaragoza, and CIBER de Enfermedades Respiratorias (CIBERES), Spain (J M Marin MD); Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland (M A Puhan MD); Hospital Universitario Araba, Sede Txagorritxu, Vitoria, Spain (P Sobradillo MD); Section of Social Medicine, Department of Public Health, Copenhagen University, Copenhagen City Heart Study, Frederiksberg Hospital, Frederiksberg, Copenhagen, Denmark (P Lange MD); Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA (A L Sternberg ScM); Centre for Research in Enviromental Epidemiology (CREAL), CIBER Epidemiología y Salud Pública (CIBERESP), Universitat Pompeu Fabra, Departament de Ciències Experimentals i de la Salut, Barcelona, Spain (J Garcia-Aymerich MD); Centre for Translational Inflammation Research, QEHB, Mindelsohn Way, Birmingham, UK (A M Turner MD); University of Michigan, Ann Arbor, MI, USA (M K Han MD); Department of Public Health and General Practice, Norwegian University of Science and Technology, Trondheim, Norway (A Langhammer MD); Centre for Clinical Documentation and Evaluation, Northern Norway Regional Health Authority, Tromso, Norway (L Leivseth PhD); Department of Clinical Science, University of Bergen, Bergen, Norway (P Bakke MD); Centre for Clinical Research, Haukeland University Hospital, Bergen, Norway (A Johannessen MD); Service de Pneumologie AP-HP, Hopitaux Universitaires Paris Centre, France (N Roche MD); James Hogg Research Centre, University of British Columbia, Vancouver, BC, Canada (D D Sin MD); and Division of Respiratory Medicine, Department of Medicine, St Paul's Hospital, Vancouver, BC, Canada (D D Sin)

Research in context

Evidence before this study

COPD affects many people worldwide, but its definition and staging remain controversial. Before 2011, the GOLD classification scheme was based exclusively on forced expiratory volume in 1 s (FEV_1) thresholds. The 2011 GOLD document proposed a new classification scheme based on spirometry, the history of exacerbations, and symptoms. Since the publication of that document, several reports have compared it with the old 2007 grading, with somewhat conflicting results.

Added value of this study

We obtained and pooled individualised data from 22 published COPD cohorts, totalling 15 632 patients that contributed 70 184 person-years of follow-up to the study. The most important finding was that the 2011 GOLD classification scheme did not improve the prediction of mortality, as compared to the 2007 classification. Through the large size of the sample, we were able to overcome many of the limitations of individual studies, and had a large statistical power for most subanalyses. Therefore, we conclude that the more complicated GOLD 2011 classification scheme is no better than the simpler previous one based on spirometry only. The use of the 2011 staging system resulted in many more patients with the most severe disease (more in GOLD D than in GOLD IV), making them seem more ill. Further, GOLD class C might be superfluous because patients in this category have a similar mortality as those in class B and treatment strategies do not differ between the two groups.

Implications of all the available evidence

Neither GOLD COPD classification schemes had sufficient discriminatory power to be used clinically for risk-classification at the individual level to predict total mortality for 3 years of follow-up and up to 10 years. It is yet to be established if an increased intensity of treatment of patients with COPD by their GOLD 2011 reclassification improves their health outcomes.

cohorts from seven countries and compared the prognostic power of the 2011 versus 2007 GOLD classification schemes to predict mortality.[6]

Methods

Study population

We searched reports published from Jan 1, 2008, to Dec 31, 2014, and obtained 28 reports as per our research protocol (appendix). Members of the COPD Cohorts Collaborative International Assessment [3CIA] Steering Committee approached the coordinators of known large, published, prospective cohort studies to gather information about their willingness to participate and the availability of a minimum required set of individual data comprising vital status (up to death, right truncation, or June 2013), age, sex, prebronchodilator and postbronchodilator FEV_1 and dyspnoea MRC grade. Among 28 identified cohorts, 22 agreed to participate and are included in our analysis.[7–28]

Details of each individual study have been published previously. Briefly, all 22 were prospective cohort studies that systematically recruited patients with COPD in publicly-funded hospitals, except for the Copenhagen City Heart Study (CCHS) and the HUNT study, which were population-based studies. All 22 studies assessed total mortality during follow-up and defined COPD by spirometric values (FEV_1 to forced vital capacity ratio of ≤0·70). All included studies were of high quality (published in peer reviewed journals), and a history of severe exacerbation (emergency room visit or hospital admission due to COPD) in the previous year. 15 cohorts contained data on severe (hospitalised) exacerbations. In the other seven studies, allocation to GOLD 2011 grades C and D was based on spirometry only. Most COPD cohorts excluded patients with a history of asthma, either self-reported or doctor-diagnosed, but required no tests of bronchial hyper-reactivity. Spirometry had to be done before and after bronchodilator administration and according to the guidelines of the American Thoracic Society/European Respiratory Society (ATS/ERS) consensus.[29]

Individual data were provided to us (3CIA) by the primary investigators of each of the participating cohorts for pooled analysis (protocol available upon request). All participants within their respective cohorts gave their informed written consent to participate in the original study, and each study was approved by the respective ethics committee.

Outcomes

We compared patients and classified them according to both 2007 and 2011 staging systems for the outcome time to death according to covariates. The primary outcome was prediction of all-cause mortality in individuals by the two systems. Secondary outcomes were prediction of all-cause mortality in subgroups of patients by gender, age bands, and smoking status, and the thresholds of variables included in the current GOLD update proposal.

Statistical analysis

We quality controlled all data centrally and created a clean dataset with a data dictionary. We queried all implausible or missing variables with the original study investigators and removed the datum from the central dataset if errors could not be corrected. Because the cohorts had different follow-up times, patients were right-censored from the date of last known follow-up.

With use of raw data, we recalculated all participant assignments to GOLD 2007 I to IV classes, and GOLD 2011 A to D stages.

We used standard semi-parametric proportional hazards Cox models to model the time-to-death data. To

	A1ATD[28]	Barmelweid[8]	Basque[11]	COPD-Gene[9]	Copenhagen City Heart Study[10]	Galdakao[12]	Gen-KOLS[13]	HUNT[14]	ICE COLD ERIC Study[15]	Initiatives IBPCO[16]	PAC-COPD Study[17]	Pamplona[5]	Requena I[18,19]	Requena II[20]	Sepoc[8]	Sevilla[21]	Terrassa I[23]	Terrassa II[24]	Terrassa III[25]	Tenerife[22]	Zaragoza I[26]	Zaragoza II[26]	Total
Number of patients	309	323	106	4484	2287	543	954	1571	409	987	342	190	174	186	318	596	135	66	181	275	137	1150	15632
Person-years	893	594	299	13858	6618	2305	7516	19182	781	2713	980	796	561	557	871	1568	584	302	704	1271	711	6520	70184
Age (years)	57·9 (9·9)	72·2 (9·1)	70·5 (8·9)	62·6 (8·6)	60·7 (9·4)	68·3 (8·3)	65·0 (10·1)	63·2 (12·9)	66·8 (9·9)	64·1 (10·5)	67·9 (8·6)	65·2 (8·5)	72·2 (8·9)	70·9 (9·0)	65·2 (9·2)	65·8 (9·6)	72·3 (9·2)	71·8 (9·0)	72·0 (9·8)	62·9 (9·9)	65·8 (7·6)	63·4 (9·4)	63·9 (10·1)
Men	185 (60%)	139 (60%)	104 (98%)	2509 (56%)	1235 (54%)	522 (96%)	583 (61%)	977 (62%)	233 (57%)	761 (77%)	318 (93%)	159 (83%)	173 (99%)	185 (99%)	318 (100%)	564 (94%)	124 (91%)	65 (99%)	172 (95%)	217 (79%)	136 (99%)	1072 (93%)	10751 (69%)
Smoking status																							
Former	63%	66%	77%	57%	19%	76%	53%	34%	55%	67%	64%	63%	75%	81%	47%	76%	71%	86%	73%	51%	72%	64%	53%
Current	4%	18%	23%	43%	71%	21%	47%	46%	41%	28%	32%	37%	22%	17%	31%	24%	17%	14%	23%	38%	27%	33%	41%
Never	15%	0	0	0	9%	3%	0	18%	5%	3%	1%	0	2%	2%	4%	0	10%	0	2%	0	0	0	4%
Missing	18%	17%	0	0	<1%	0	0	2%	<1%	2%	3%	0	1%	0	19%	0	2%	0	3%	12%	1%	2%	2%
BMI (kg/m^2)	25·7 (4·9)	26·0 (6·3)	26·1 (4·9)	27·9 (6·1)	25·0 (4·2)	28·3 (4·4)	25·4 (4·9)	26·4 (4·4)	26·2 (5·2)	25·6 (6·0)	28·2 (4·7)	27·0 (4·4)	28·0 (4·2)	28·1 (5·2)	26·4 (4·2)	29·2 (5·7)	26·3 (4·9)	25·7 (4·3)	27·9 (5·0)	27·3 (5·1)	27·7 (4·6)	27·5 (4·8)	26·9 (5·4)
FEV_1 (% pred)	53·1 (25·1)	45·2 (16·1)	46·9 (11·4)	57·4 (22·8)	70·5 (23·7)	55·0 (13·3)	46·9 (17·0)	63·8 (18·7)	55·5 (16·6)	53·2 (20·2)	52·4 (16·2)	68·9 (19·9)	48·1 (16·8)	44·5 (16·5)	45·0 (18·3)	43·5 (13·3)	41·3 (13·0)	30·2 (12·9)	45·2 (14·4)	55·8 (21·2)	49·8 (17·6)	62·3 (20·3)	57·5 (21·9)
GOLD 2007																							
Mild (I)	16%	4%	1%	18%	34%	<1%	1%	19%	2%	9%	6%	34%	3%	5%	0·6	<1%	0	2%	0	15%	4%	21%	16%
Moderate (II)	29%	33%	38%	43%	46%	66%	47%	58%	63%	45%	48%	48%	36%	25%	39%	34%	23%	3%	35%	41%	43%	51%	46%
Severe (III)	40%	45%	58%	26%	15%	31%	34%	19%	25%	32%	39%	15%	33%	54%	35%	49%	59%	32%	52%	35%	37%	24%	28%
Very severe (IV)	15%	18%	4%	14%	4%	3%	19%	3%	9%	14%	8%	3%	14%	15%	25%	17%	19%	58%	12%	10%	15%	5%	11%
Missing	0	0	0	0	<1%	0	0	0·5	0	<1%	0	0	14%	1%	0	0	0·0	6%	1%	0	1%	0	<1%
mMRC dyspnoea score	2·3 (1·2)	1·4 (1·1)	2·0 (0·6)	1·9 (1·5)	1·1 (1·3)	1·4 (0·9)	1·4 (1·3)	1·7 (1·3)	1·9 (1·5)	1·7 (1·1)	1·2 (1·0)	1·0 (1·1)	1·8 (1·2)	1·1 (1·5)	2·1 (1·5)	1·4 (1·0)	2·4 (1·3)	2·0 (1·0)	2·8 (1·2)	2·1 (1·2)	1·7 (1·1)	1·7 (1·1)	1·7 (1·3)
Past year exacerbation																							
Yes	52%	0	0	16%	0	0	15%	0	10%	39%	4%	0	3%	16%	0	25%	25%	39%	30%	6%	0	17%	54%
No	49%	0	0	84%	0	26%	85%	0	91%	59%	96%	0	97%	80%	0	75%	75%	55%	69%	86%	0	81%	13%
Missing	0	100%	100%	0	100%	74%	0	100%	0	2%	0	100%	0	5%	100%	0	0	6%	1%	8%	100%	2%	34%

Data are n, %, or mean (SD), unless otherwise stated. BMI=body-mass index. FEV_1=forced expiratory volume in 1 s. % pred=percentage predicted. GOLD=Global Initiative for Chronic Obstructive Lung Disease. mMRC=modified Medical Research Council.

Table 1: **Demographic and clinical characteristics of patients at baseline or enrolment, by cohort**

Correspondence to:
Dr Joan B Soriano, Instituto de Investigación Hospital Universitario de la Princesa (IISP), Universidad Autónoma de Madrid, Cátedra UAM-Linde, Madrid 28030, Spain
jbsoriano2@gmail.com

assess the relation between FEV_1 as percent predicted and total mortality without imposing any a priori shape, we used penalised splines method, a type of non-parametric regression, based on the Akaike information criteria (AIC) in which all models were stratified by the cohort and with FEV_1=100% predicted as the referent (HR=1).[30] Splines enable fitting of regression models to data without any parametric restrictions. We used area under incident-cumulative receiver operating characteristic (ROC) curve (AUC) to assess the performance of the time-to-event models.[31] Heterogeneity and potential study effects were evaluated with the individual study as an additional stratification variable. Statistical comparisons of curves and AUCs were made by employing 10 000 iterations of the general bootstrap algorithm (gBA).[32]

We formally assessed heterogeneity between studies when comparing the GOLD 2007 and 2011 staging. Assuming that the maximum heterogeneity would occur when patients with COPD would be equally distributed

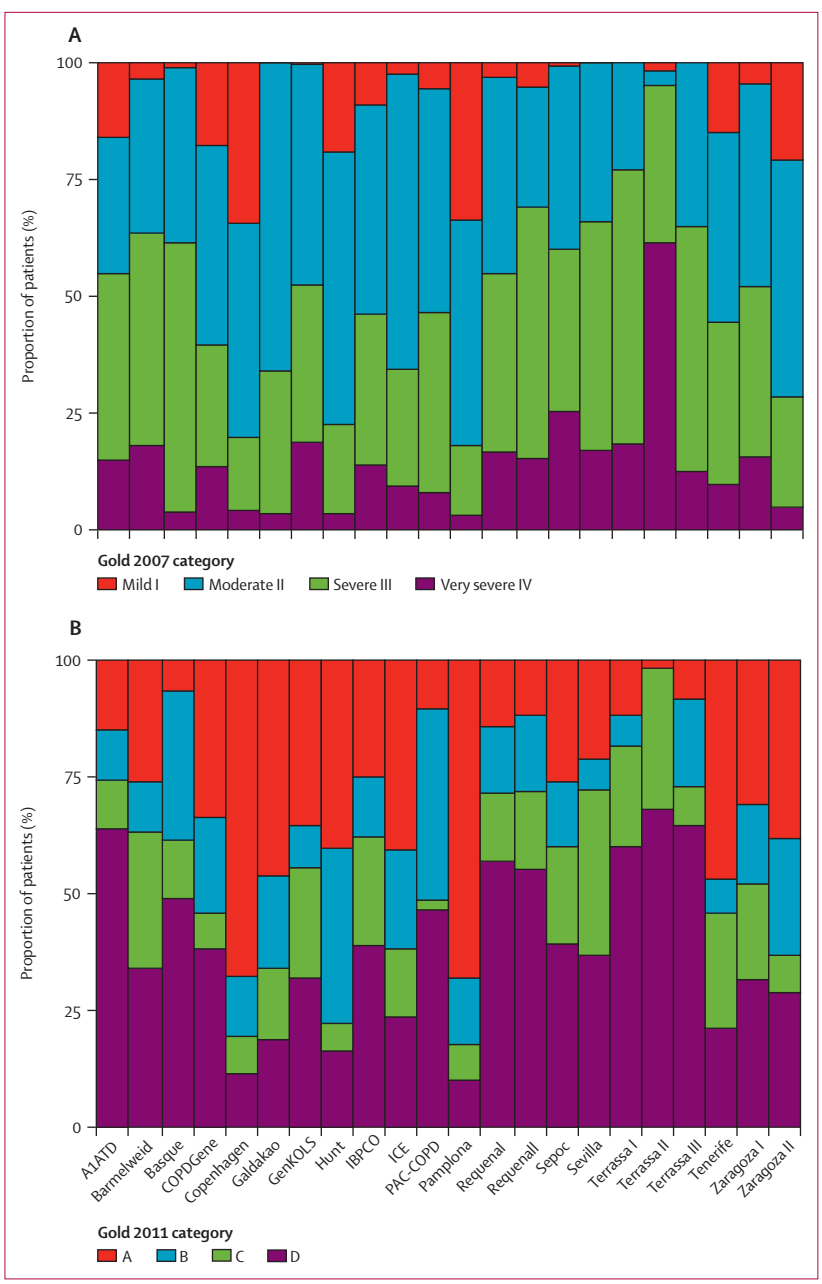


Figure 1: **Distribution of participants by classification in GOLD 2007 (A) and GOLD 2011 (B)**
n=15 632. GOLD=Global Initiative for Chronic Obstructive Lung Disease.

	Men (n=10 751)	Women (n=4881)	Total (n=15 632)
Mortality of study population	60·6	33·9	51·7
Mortality by age			
≤65 years	33·5	20·3	28·6
>65 years	91·1	57·6	81·9
Mortality by GOLD 2007 status			
Mild (I)	25·5	13·8	21·1
Moderate (II)	45·6	26·9	39·3
Severe (III)	83·5	46·8	73·2
Very severe (IV)	130·8	96·8	121·1
Mortality by mMRC dyspnoea (0–4) score			
<2	41·1	19·6	34·5
≥2	86·5	46·2	71·8
Mortality by exacerbations			
<2	56·6	30·5	49·0
≥2	99·2	51·2	83·5
Mortality by GOLD 2011 status			
Grade A	31·9	15·7	26·7
Grade B	56·5	31·3	45·9
Grade C	67·9	35·1	59·5
Grade D	107·2	62·4	93·0
Mortality by smoker status			
Former	69·1	35·5	60·3
Current	50·8	31·1	43·6
Never	36·6	36·3	36·4
Mortality by cohort			
ADO	50·5	26·2	41·7
COCOMICS	83·3	28·2	79·3
Others	50·4	35·7	44·4
Mortality by country			
Denmark	49·4	32·0	41·9
Spain	79·7	31·1	75·6
UK	5·5	11·3	7·8
Switzerland	112·1	119·9	114·3
USA	39·4	30·3	35·2
Norway	54·7	37·0	47·5
France	57·5	37·0	53·2

GOLD=Global Initiative for Chronic Obstructive Lung Disease. mMRC=modified Medical Research Council.

Table 2: **Mortality rates per 1000 person-years by sex**

in the four groups (ie, each of the groups contains 25% of patients), and that the minimum heterogeneity would be when a single category includes 100% of patients, the H index of heterogeneity was calculated, where P_i is the proportion of individuals in the i[th] group.

$$H = 1 - \frac{2}{3}\sum_{i=1}^{4} | p_i - 1/4 |$$

This H index ranges between 0 and 1, where 0 is minimal heterogeneity (all individuals are concentrated in one category) and 1 is maximum heterogeneity (each group has 25% of individuals). We also calculated their 95% CI. We considered p values of less than 0·05 to be statistically significant.

For all analyses we used software R. We used packages risksetROC and survival, both available in the Comprehensive R Archive Network. Further explanatory text and references are in the appendix.

Role of the funding source

We received no funding for this work. JBS accepts final responsibility for the integrity of the work as a whole, and had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the analysis.

Results

We obtained and pooled data for individuals from 22 published COPD cohorts (table 1). 15 632 patients contributed 70 184 person-years of follow-up time to the study.

Based on the 2007 GOLD classification scheme, of 15 882 patients with applicable data, 2424 (16%) had mild (stage I), 7142 (46%) had moderate (stage II), 4346 (28%) had severe (stage III), and 1670 (11%) had very severe (IV) disease (table 1). Based on the GOLD 2011 scheme, of 14 660 patients with applicable data: 5548 (38%) were grade A, 2733 (19%) were grade B, 1835 (13%) were grade C, and 4544 (31%) were grade D.

We compared how the 14 660 patients were classified with GOLD 2007 and GOLD 2011 classifications. Concordance of the two schemes was reasonable in the mild categories with 1788 (78%) of 2279 patients with stage I in GOLD 2007 being reclassified as stage A in the GOLD 2011 (appendix). However, most patients in the 2007 GOLD system were categorised as moderate (6694 [46%]) and severe (4085 [28%]; appendix) whereas most were class D (very severe; 4544 [31%]) in the 2011 scheme (figure 1), resulting in nearly three times more COPD patients in stage D than in former stage IV (p<0·05).

There was significant heterogeneity in the distribution of patients across individual studies irrespective of the classification scheme (figure 1). Overall GOLD 2011 made the COPD patient distribution more heterogeneous than GOLD 2007; the H heterogeneity index for GOLD 2007 was 0·684 (95% CI 0·674–0·693), with 10 779 (74%) patients being in the intermediate groups whereas for GOLD 2011 it was 0·749 (95% CI 0·739–0·758) with 10 092 individuals (69%) in the two most common groups, A and D. These differences are significant because their 95% CIs do not overlap.

Overall, men had a higher mortality rate than had women (table 2). Similarly, patients with an mMRC dyspnoea score of two or higher had a two times higher risk of death than had those with a score of zero or one. A history of exacerbation was also a significant risk factor for mortality with two or more severe exacerbations

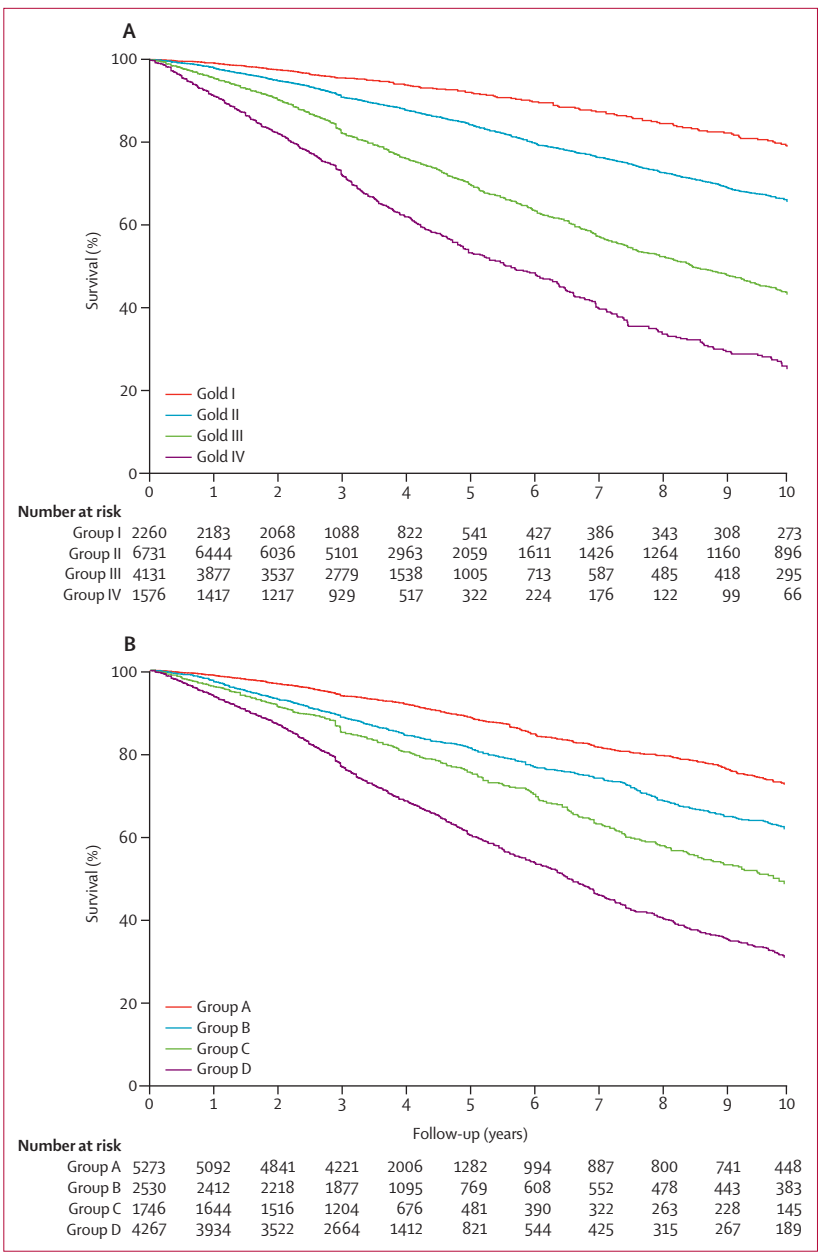
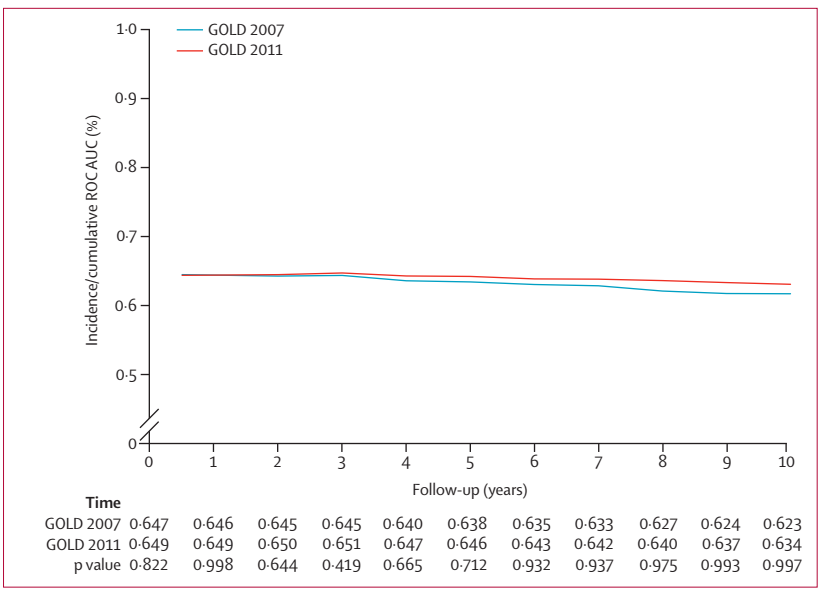


Figure 2: **Kaplan-Meier survival curves by GOLD 2007 (A) and GOLD 2011 (B)**
n=15 632. GOLD=Global Initiative for Chronic Obstructive Lung Disease.

For more on **software R** see www.r-project.org

See **Online** for appendix



Time	0	1	2	3	4	5	6	7	8	9	10
GOLD 2007	0·647	0·646	0·645	0·645	0·640	0·638	0·635	0·633	0·627	0·624	0·623
GOLD 2011	0·649	0·649	0·650	0·651	0·647	0·646	0·643	0·642	0·640	0·637	0·634
p value	0·822	0·998	0·644	0·419	0·665	0·712	0·932	0·937	0·975	0·993	0·997

Figure 3: **Change by time of area under the incidence/cumulative ROC curve to predict death up to 10 years of GOLD 2007 and GOLD 2011**
n=15 632. GOLD=Global Initiative for Chronic Obstructive Lung Disease. ROC=receiver operating characteristic. AUC=area under the curve.

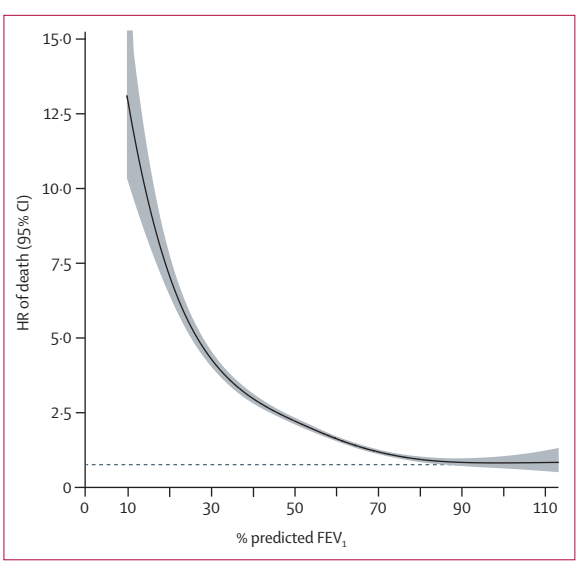


Figure 4: **Spline of the HR of death to identify spirometry thresholds of severity**
n=15 632. 100% predicted FEV_1 is taken as the reference point (hazard ratio [HR]=1). FEV_1=forced expiratory volume in 1 s.

conferring a risk that was 1·7 times higher than those with zero or one exacerbations in the previous year (all p<0·05). According to the 2007 scheme, there was a severity-dependent relationship between GOLD categories and total mortality, with very severe disease conferring a relative risk that was 5·7 times higher than that of mild disease; moderate and severe disease increased the risk by 1·9 and 3·5 times, respectively (table 2). In the GOLD 2011 classification scheme, patients with class D severity had the worst prognosis with a risk 3·5 times that of patients with class A severity. Interestingly, risk of mortality did not significantly differ between classes B and C, which increased the risk by 1·7–2·2 times compared with class A. There were higher mortality rates in the Swiss cohort, whereas in the UK cohort there were only seven deaths in total (table 2).

The comparison of mortality data between 2007 and 2011 GOLD classification schemes is depicted in figure 2. Although the GOLD 2007 classification scheme contained fewer variables, one versus three in GOLD 2011, both schemes were equivalent in prediction of death in the short term and long term (figure 3). The predictive capacity for survival up to 10 years was significant for both systems (p<0·01) but area under the curves were only 0·623 (GOLD 2007) and 0·634 (GOLD 2011), and GOLD 2007 and 2011 did not differ significantly by predictive power. Detailed analyses of mortality stratified by demographic factors are in the appendix. The discriminatory power of GOLD 2011 by sex was significantly better in men (distance 0·779) than in women (distance 0·616; p=0·039; appendix). However, this pattern was not seen by age or smoking status (appendix).

Finally, we established the optimal cutoff for FEV_1 thresholds based on ROC-AUC for total mortality in the dataset (figure 4). With the spline analysis, the highest ROC-AUC values were obtained by using % predicted FEV_1 thresholds of 85%, 55%, and 35% to predict total mortality, which are very similar to the ones used in the GOLD stratification (80%, 50%, and 30%). A post-hoc analysis of these cutoff values stratified by MRC dyspnoea score, age, COPD exacerbation, and sex confirmed the initial results (appendix).

Discussion

This is the largest study to evaluate total mortality in patients with COPD according to severity with more than 15 000 patients and 70 000 person-years of follow-up. The most important finding was that the use of the 2007 GOLD classification scheme (based exclusively on FEV_1) and the use of the 2011 GOLD classification scheme (based on spirometry, a history of exacerbations, and symptoms) similarly predicted mortality within 10 years. However, although both classification schemes were significantly related to total mortality up to 10 years, neither had a striking discriminatory power with AUCs ranging from 0·62 to 0·65. Thus, these classification schemes might be useful to guide therapy but are unlikely to be clinically useful to identify patients at high risk of mortality, even in the short term.

Another important finding was that in the 2011 GOLD classification scheme, we found little difference between categorisation in classes B and C within the first 3 years of follow-up to predict mortality, as previously noted.[5,33] Our data provide additional support that class C might be superfluous given that, within this pooled analysis, fewer than 1835 of 15 632 (12%) of patients with COPD fit into

this category (appendix) and little difference exists in the treatment strategies between classes B and C.[3] Moreover, the 2011 GOLD classification scheme shifts more of the patient population to the very severe category (class D), nearly three times more than in the GOLD 2007 stage IV, making the patients seem more ill than in the 2007 staging system. The heterogeneity index was significantly higher in GOLD 2011 than in GOLD 2007, given the overall shift to the most severe stage D for many COPD patients. This shift might lead to increased intensity of treatment of patients with COPD.[2,3,33] In both the 2007 and 2011 schemes, FEV_1-based severity was categorised based on cutoffs of 80%, 50%, and 30%.[3,6] In this pooled dataset, the most optimum thresholds were very similar (85%, 55%, and 35%) but only 2956 of 15 582 (19%) participants would change spirometric severity in the four category stages (mild, moderate, severe, and very severe) and only 6% when using the lower than 50% threshold in GOLD 2011 (appendix). This supports sustaining the current cutoffs of the GOLD spirometric grades, which are also recommended by the ATS/ERS COPD guidelines.

By pooling individualised data, we were able to reduce the errors in data analysis and interpretation that plague many large-scale meta-analyses.[34] Further, although each of the studies included in this dataset had unique features, there were commonalities in data collection (eg, collection of spirometry and dyspnoea and mortality) that enabled direct comparisons of the 2007 and 2011 GOLD classification schemes.

Beyond respiratory semantics and historical controversies in pneumology, the COPD GOLD gradings are often used in patients with α-1 antitrypsin deficiency (AATD). Using GOLD 2011, investigators successfully identified patients with increased risk of poorer outcomes in AATD,[28] and further studies to identify subgroups of patients most likely to benefit from augmentation therapy are indicated.

There were some limitations to our analysis. First, although we were able to include 22 individual studies, six large COPD cohort studies were not included in our analysis. There were statistically significant differences in age, sex, and smoking status when comparing cohorts included by the 3CIA and those that did not participate (appendix), mostly due to the large sample sizes. However, there is no compelling reason to believe that the non-inclusion of these studies distorted our principal findings. Second, there was significant variability in the severity and outcomes across the individual studies included. However, this might also be considered a strength because it enables inclusion of patients across the full range of COPD disease severity, age, smoking status, and other significant drivers of mortality. Third, we did not compare the GOLD classification scheme with multidimensional indices of COPD (eg, BODE and other indices) because this was beyond the scope of this analysis and more importantly because previous studies have clearly shown the superiority of multidimensional indices over spirometry-based classification schemes in prediction of mortality in patients with COPD.[5,8] Regretfully, we realised definitions for specific comorbidities and cause of death within each of the 22 cohorts were too heterogeneous to be analysed all together, and subject to bias and error, and so discarded these analyses. Finally, the spline statistical technique is useful in numerical continuous variables, but it cannot be used for discrete variables such as mMRC dyspnoea, where standard Cox methods indicate that an mMRC threshold of two or higher is the best discriminant of mortality in COPD patients (appendix).

The 2007 GOLD classification scheme based exclusively on spirometry predicts mortality (up to 10 years) equally well as the 2011 GOLD classification scheme. Although the inclusion of patient symptoms and exacerbation history might guide therapy, they add very little to prognosis. Importantly, neither classification scheme has sufficient discriminatory power to be used clinically for risk classification at the individual level. The use of the 2011 staging system resulted in many more patients with the most severe disease (three-fold more in GOLD D in 2011 than in GOLD IV in 2007). This change has the potential to increase therapeutic intensity in these patients, which might have important economic and clinical impact. It remains to be seen whether this change will effectively improve health outcomes.

Contributors

The 3CIA protocol was developed by JBS, BL and DDS, with assistance from ASR and PM-C. Statistical analysis was conducted by PM-C and BK. The initial manuscript was drafted by JBS, BL, DDS, ASR, and PM-C. All other authors collected data and led their individual cohorts and respective publications, and helped to write this manuscript and agreed with the decisions made about it.

Declaration of interests

We declare no competing interests.

Acknowledgments

None of the funding sources of the individual trials were involved in any aspect of the 3CIA initiative, including the design, data collection and analysis, decision to publish, or preparation of the manuscript. PM-C was supported by research grant MTM2011-23204 from the Spanish Ministerio de Ciencia e Innovación (FEDER support included). JG-A has a researcher contract from the Instituto de Salud Carlos III (CP05/00118), Ministry of Health, Spain. We thank the peer reviewers of *The Lancet Respiratory Medicine* for their constructive assessment and suggestions, which significantly increased the quality and scope of our original submission.

References

1 Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; **380:** 2095–28.

2 Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS, for the GOLD Scientific Committee. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *Am J Respir Crit Care Med* 2001; **163:** 1256–76.

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:** 347–65.

4 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:** 636–46.

5 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.

6 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:** 532–55.

7 Almagro P, Martinez-Camblor P, Soriano JB, et al. Finding the best thresholds of FEV1 and dyspnea to predict 5-year survival in COPD patients: the COCOMICS study. *PLoS One* 2014; **9:** e89866.

8 Puhan MA, Garcia-Aymerich J, Frey M, et al. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet* 2009; **374:** 704–11.

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

10 Vestbo J, Prescott E, Lange P. Association of chronic mucus hypersecretion with FEV_1 decline and chronic obstructive pulmonary disease morbidity. Copenhagen City Heart Study Group. *Am J Respir Crit Care Med* 1996; **153:** 1530–35.

11 Puhan MA, Hansel NN, Sobradillo P, et al. Large-scale international validation of the ADO index in subjects with COPD: an individual subject data analysis of 10 cohorts. *BMJ Open* 2012; **2:** e002152.

12 Esteban C, Quintana JM, Aburto M, et al. The health, activity, dyspnea, obstruction, age, and hospitalization: prognostic score for stable COPD patients. *Respir Med* 2011; **105:** 1662–70.

13 Johannessen A, Nilsen RM, Storebo M, Gulsvik A, Eagan T, Bakke P. 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.

14 Leivseth L, Brumpton BM, Nilsen TI, Mai XM, Johnsen R, Langhammer A. GOLD classifications and mortality in chronic obstructive pulmonary disease: the HUNT Study, Norway. *Thorax* 2013; **68:** 914–21

15 Frei A, Muggensturm P, Putcha N, et al. Five comorbidities reflected the health status in patients with chronic obstructive pulmonary disease: the newly developed COMCOLD index. *J Clin Epidemiol* 2014; **67:** 90–111.

16 Roche N, Deslee G, Caillaud D, et al. Impact of gender on COPD expression in a real-life cohort. *Respir Res* 2014; **15:** 20.

17 Garcia-Aymerich J, Gomez FP, Benet M, et al. Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes. *Thorax* 2011; **66:** 430–37.

18 Soler JJ, Sanchez L, Roman P, Martinez MA, Perpina M. Risk factors of emergency care and admissions in COPD patients with high consumption of health resources. *Respir Med* 2004; **98:** 318–29.

19 Soler-Cataluña JJ, Martinez-Garcia MA, Roman Sanchez P, Salcedo E, Navarro M, Ochando R. Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease. *Thorax* 2005; **60:** 925–31.

20 Soler-Cataluña JJ, Martinez-Garcia MA, Sanchez LS, Tordera MP, Sanchez PR. Severe exacerbations and BODE index: two independent risk factors for death in male COPD patients. *Respir Med* 2009; **103:** 692–99.

21 Alfageme I, Reyes N, Merino M, et al. The effect of airflow limitation on the cause of death in patients with COPD. *Chron Respir Dis* 2010; **7:** 135–45.

22 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:** 591–97.

23 Almagro P, Calbo E, Ochoa de Echaguen A, et al. Mortality after hospitalization for COPD. *Chest* 2002; **121:** 1441–48.

24 Sanjaume M, Almagro P, Rodriguez-Carballeira M, Barreiro B, Heredia JL, Garau J. Post-hospital mortality in patients re-admitted due to COPD. Utility of BODE index. *Rev Clin Esp* 2009; **209:** 364–70 (in Spanish).

25 Almagro P, Salvado M, Garcia-Vidal C, et al. Recent improvement in long-term survival after a COPD hospitalisation. *Thorax* 2010; **65:** 298–302.

26 Marin JM, Soriano JB, Carrizo SJ, Boldova A, Celli BR. Outcomes in patients with chronic obstructive pulmonary disease and obstructive sleep apnea: the overlap syndrome. *Am J Respir Crit Care Med* 2010; **182:** 325–31.

27 Fishman A, Martinez F, Naunheim K, et al. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* 2003; **348:** 2059–73.

28 Pillai AP, Turner AM, Stockley RA. Global Initiative for Chronic Obstructive Lung Disease 2011 symptom/risk assessment in alpha1-antitrypsin deficiency. *Chest* 2013; **144:** 1152–62.

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

30 Malloy EJ, Spiegelman D, Eisen EA. Comparing measures of model selection for penalized splines in Cox models. *Comput Stat Data Anal* 2009; **53:** 2605–16.

31 Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics* 2005; **61:** 92–105.

32 Martínez-Camblor P, Corral N. A general bootstrap algorithm for hypothesis testing. *J of Stat Plann Infer* 2012; **142:** 589–600.

33 Soriano JB. The GOLD Rush. *Thorax* 2013; **68:** 902–03.

34 Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010; **340:** c221.