

Adjusting for COPD severity in database research: developing and validating an algorithm

Lucas MA Goossens¹
Christine L Baker²
Brigitta U Monz³
Kelly H Zou²
Maureen PMH Rutten-van
Mölken¹

¹Institute for Medical Technology
Assessment, Erasmus University,
Rotterdam, The Netherlands;

²Pfizer Inc, New York City, NY, USA;

³Boehringer Ingelheim International
GmbH, Ingelheim am Rhein, Germany

Purpose: When comparing chronic obstructive lung disease (COPD) interventions in database research, it is important to adjust for severity. Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines grade severity according to lung function. Most databases lack data on lung function. Previous database research has approximated COPD severity using demographics and healthcare utilization. This study aims to derive an algorithm for COPD severity using baseline data from a large respiratory trial (UPLIFT).

Methods: Partial proportional odds logit models were developed for probabilities of being in GOLD stages II, III and IV. Concordance between predicted and observed stage was assessed using kappa-statistics. Models were estimated in a random selection of 2/3 of patients and validated in the remainder. The analysis was repeated in a subsample with a balanced distribution across severity stages. Univariate associations of COPD severity with the covariates were tested as well.

Results: More severe COPD was associated with being male and younger, having quit smoking, lower BMI, osteoporosis, hospitalizations, using certain medications, and oxygen. After adjusting for these variables, co-morbidities, previous healthcare resource use (eg, emergency room, hospitalizations) and inhaled corticosteroids, xanthines, or mucolytics were no longer independently associated with COPD severity, although they were in univariate tests. The concordance was poor (kappa = 0.151) and only slightly better in the balanced sample (kappa = 0.215).

Conclusion: COPD severity cannot be reliably predicted from demographics and healthcare use. This limitation should be considered when interpreting findings from database studies, and additional research should explore other methods to account for COPD severity.

Keywords: GOLD, healthcare resource use, partial proportional odds logit

Introduction

Treatment effectiveness and healthcare resource use in daily practice may be different from effectiveness and resource use in clinical trials. This is increasingly recognized by researchers, policy makers and decision makers responsible for pricing and reimbursement of healthcare interventions. Hence, the importance of data generated from sources that reflect the use and the associated outcomes in routine practice settings is growing. Suitable sources of such real-life data could be health records kept by physicians (eg, their routine records or specifically established databases; paper-based or electronic), patient registries that enroll patients with specific diseases or other characteristics of interest (eg, cancer registries), or administrative claims databases of healthcare insurers and provider organizations set up for the purpose of reimbursement of providers for their expenses.

Correspondence: Lucas Goossens
Institute for Medical Technology
Assessment, Erasmus University,
PO Box 1738, 3000 DR Rotterdam,
The Netherlands
Tel +31 10 4088582
Fax +31 10 4089081
Email goossens@bmg.eur.nl

The inherent problem of analyses conducted on such databases relates to the fact that the data are often collected for other purposes. When treatment is not assigned to patients at random, disease severity and prognosis of patients may differ systematically across treatments. In order to adjust for this, data that reflect disease severity must be available.

The currently most widely used severity classification is the spirometric classification of the Global Initiative for Chronic Obstructive Lung Disease (GOLD),¹ which is based on lung function in terms of the forced expiratory volume in one second as a percentage of the value predicted for sex, age and height (FEV_1 -%predicted). It is used to diagnose COPD, to monitor disease progression and to aid in treatment decisions. Lung function is an important indicator of COPD severity because it is associated with mortality,² exacerbations,^{3,4} health care utilization and costs.⁵ There is also an association between lung function and quality of life⁶ although this relationship is not as strong.^{7,8} Consequently, the same associations have been shown for the GOLD classification as it is based on FEV_1 .⁹⁻¹³

It is widely recognized, also by GOLD,¹ that the impact of COPD does not only depend on lung function and furthermore that is not only a lung disease but a multi-dimensional disease with many systemic, extra-pulmonary effects. Therefore, composite measures of COPD severity have been proposed. However, almost all of these still accept the importance of airflow limitation and they are partly based on the FEV_1 -%predicted.¹⁴⁻¹⁸

However, lung function measures obtained through spirometry are not routinely available in databases and registries. Most routine databases also lack information on the other parameters that constitute the more recent composite measures of COPD severity. Previous retrospective database research has approximated COPD severity using demographic (eg, age and/or smoking status) and resource utilization data (eg, medications used and/or hospital admission).¹⁹⁻²⁴

In this study, we attempted to develop a multivariable predictive algorithm to derive the severity of COPD as classified by GOLD, using variables that are commonly available in routine databases. The analysis was performed on the baseline data of the “Understanding Potential Long-term Impacts on Function with Tiotropium” (UPLIFT) trial, a large, 4-year trial in COPD patients with rate of decline in FEV_1 as primary endpoint.²⁵ This trial was especially suitable for this purpose because of its size (with almost 6000 patients randomized), and because among the baseline data collected are many data commonly available in routine databases (eg, demographics, medications used in the past, hospital admissions in the year

preceding enrollment). Moreover, FEV_1 was collected with high quality, thus providing confidence in the GOLD severity assignment of trial patients.

Material and methods

Data source

We used baseline data from the UPLIFT trial. The UPLIFT trial was a randomized, double-blind, placebo-controlled 4-year trial, investigating the effect of tiotropium on the yearly rate of decline in FEV_1 in patients with moderate to very severe COPD according to GOLD (stages II to IV). Patients were permitted to use all respiratory medications except inhaled anticholinergics. 5993 patients were randomized.²⁵ Main inclusion criteria besides a diagnosis of COPD included age of 40 years or over, a smoking history of at least 10 pack-years, a post-bronchodilator FEV_1 -%predicted of 70% or less, and an FEV_1 of 70% or less of the forced vital capacity. Key exclusion criteria were a history of asthma, a COPD exacerbation or respiratory infection within 4 weeks before screening, use of supplemental oxygen for more than 12 hours per day. UPLIFT collected information on patients’ demographics, comorbidities and co-medications as well as COPD characteristics and smoking history, along with contacts with health care providers in various settings in the year preceding enrollment.

In order to enhance the homogeneity of treatment patterns, the sample was limited to 3698 patients from Western Europe, the United States, Australia and New Zealand with COPD severity stage II, III or IV. The sample was randomly split into two sets: one for developing the algorithm (2/3 of patients) and the other for validation (1/3 of patients).

Selection of variables

Potential variables for inclusion in the algorithm to approximate disease severity were those used in previous database research and those used in diagnostic and severity classifications.^{14-17,26-29} Variables were then selected if available in UPLIFT and likely to be available in routine databases. The final list of variables to be tested consisted of age (continuously in years, squared, and in categories: 40–49, 50–59, 60–69, 70–79, 80–90, 90+). BMI (continuously and categorized: <18.5, 18.5–21.0, 21.0–25.0, 25.0–30.0, >30.0), smoking status (current/former smoker), pack-years of smoking (continuously and categorized: <20, 20–60, >60), sex, type of respiratory maintenance medications (short-acting and long-acting bronchodilators, inhaled corticosteroids (ICS), mucolytics, leukotriene modifiers, xanthine), number of medication types (0 to 5), treatment

of exacerbations (number of courses of antibiotics or oral steroids, categorized: 0, 1, 2, ≥ 3 in one year), resource use in one year (number of emergency room visits with/without hospital admission, categorized (0, 1, ≥ 2), number of scheduled and unscheduled general practitioner (GP) visits (categorized: 0, 1, 2, 3–4, ≥ 5 and 0, 1, 2, ≥ 3 , respectively), hospital admissions (yes/no), use of oxygen at home (yes/no), statin use (yes/no), use of other cardiovascular medication (yes/no), number of co-morbidities (categorized: 0, 1, 2, 3–5, 6–9, ≥ 10), Charlson comorbidity index³⁰ (categorized: 1, 2, 3, ≥ 4), presence of selected co-morbidities (arrhythmia, coronary heart disease, vascular disease, hypertension, disorders of nervous system, stroke, diabetes, depression, anemia, and platelet disorders, osteoporosis and cataract), as well as interaction terms (age \times gender, age \times BMI, BMI \times gender).

Analysis

Several partial proportional odds (PPO) ordered logit models were used to estimate the probabilities of being in GOLD stages II (FEV₁%-predicted $> 50\%$), III (FEV₁%-predicted FEV₁%-predicted 30%–50%), and IV (FEV₁%-predicted $< 30\%$), as COPD severity categories are ordered. The “development dataset” was used for this purpose.

The PPO resembles the standard ordered logit model, which is the best-known ordered regression model. In our case, with three possible outcomes Y , it estimates the probability that $Y > \text{stage II}$ and the probability that $Y > \text{stage III}$. Ordered regression models assume that the observed ordered outcome Y is a function of a continuous and unobservable variable Y^* . Two thresholds τ_j are assumed to determine in which stage a patient is classified: $Y > Y_j$ if $Y^* > \tau_j$, where j denotes a specific outcome and τ_j is the upper limit of Y^* for this outcome. Y^* is related to the explanatory variables X : $Y_i^* = X_i\beta_j + \varepsilon_{ij}$, where i denotes individual patients. Since the latent variable Y^* does not completely equal the sum of the products of coefficients and variable values, the outcome Y cannot be determined with certainty from the data. Assuming a logit distribution for the random error term makes it possible to model the probability that a patient is in a certain stage or worse ($Y^* > \tau_j$). This probability may be written as: $\Pr(Y_i > j) = \exp(-\tau_j + X_i\beta_j) / 1 + \exp(-\tau_j + X_i\beta_j)$, $j = 1, 2, \dots, M - 1$. M is the number of possible outcomes.

In contrast to the PPO model, however, the ordered logit model can only estimate one coefficient for each explanatory variable. This coefficient is assumed to be identical for all dichotomizations of the outcome variable (in this case, stage IV versus II/III and stage III/IV versus II), and thus

coefficient $\beta_j = \beta$. This is the proportional odds or parallel regressions assumption. If this assumption is violated, which often happens in practice, estimates are invalid and important differences in the relationships at different thresholds may go unnoticed. The partial proportional odds model relaxes this assumption for variables where it does not hold.³¹ For these variables, a coefficient is estimated for each dichotomization.

The final model was developed by stepwise backward elimination from the full model, which contained all variables. In each step the variable with the largest P -value was eliminated and the model was re-estimated. This process was repeated until all variables had at least a value of $P \leq 0.10$.

We performed two sensitivity analyses in order to account for the fact that the proportion of patients with very severe COPD (GOLD IV) in the dataset was smaller than the proportion of patients with moderate or severe COPD. Firstly, the final model was re-estimated in a sub-sample with a balanced distribution of patients across all three severity stages. This balance was achieved by using all GOLD stage IV patients and random draws from patients in stages II and III. Secondly, the patients in GOLD stage III and IV were combined into one group. The probability of being in ‘severe/very severe’ was then analyzed in a binary logit model with the same variables as in the final PPO model.

All analyses were performed in Stata 11.1, (StataCorp LP, College Station, TX) using the `gologit2` command for the PPO model.³² Univariate tests were analysis of variance (ANOVA) for continuous variables and chi-square for categorical variables. Statistical significance was reached when a two-sided P -value was ≤ 0.05 .

Predictive performance

The regression results were used to predict the probabilities of being in each GOLD severity stage for each patient in the validation dataset. The predicted stage was defined as the stage with the highest predicted probability. The agreement between the predicted and the observed stage was assessed using kappa statistics.³³ For a kappa statistic, a value of 0 indicates that agreement has occurred by chance, whereas a value of 1 indicates perfect agreement. No generally accepted interpretation of the magnitude of the kappa-statistic exists. Landis and Koch suggested the following labels of agreement for ranges of values, which are often quoted: slight (up to 0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.0).³⁴ We present kappa statistics

unweighted and weighted. In the weighted kappa statistics, patients misclassified in the neighboring category (ie, II or IV instead of III, or III instead of II or IV) count as 0.5 agreement and patients misclassified in a non-neighboring category (II instead of IV or vice versa) as no agreement).

In addition to the agreement per disease stage, the overall agreement was calculated. For the binary logit model, the c-statistic was computed as a measure of its discriminative performance. The c-statistic represents the area under the receiver operating characteristic (ROC) curve, a plot of sensitivity against specificity. Values can range from 0.5 (no predictive ability) to 1 (perfect discrimination).

Lastly, in an analysis of correctly and incorrectly classified patients, the average values of FEV₁%-predicted were compared per severity stage.

Results

Description of sample

Patient characteristics are summarized in Table 1A, demonstrating that 47% were in GOLD stage II, 44% in stage III, and 9% in stage IV. Patients in more severe disease states were more likely to have a lower BMI, to have quit smoking and to suffer from osteoporosis. The mean age of patients in stage IV was 2–3 years lower than in the less severe

Table 1A Patient characteristics, risk factors, and comorbidity per GOLD stage in full sample (development and validation sets combined)

Stage	II n = 1720 (47%)	III n = 1640 (44%)	IV n = 338 (9%)	P-value*
Age (SD)	64.9 (8.47)	65.8 (7.98)	63.0 (8.23)	0.000
Male	70.10%	72.80%	75.40%	0.066
BMI < 21	9.94%	12.56%	26.63%	0.000
BMI > 25	62.15%	56.58%	34.61%	
Underweight (BMI < 18.5)	2.44%	3.84%	10.06%	0.000
Low weight (18.5 > BMI < 21)	7.50%	8.72%	16.57%	
Normal BMI	27.91%	30.85%	38.76%	
Overweight (25 < BMI < 30)	36.92%	36.45%	23.37%	
Obese (BMI > 30)	25.23%	20.12%	11.24%	
Current smoker	33.43%	28.23%	27.51%	0.002
<20 Pack years	6.63%	5.85%	7.69%	0.700
20–60 pack years	66.86%	66.95%	64.79%	
>60 pack years	26.51%	27.20%	27.51%	
Comorbidities				
0	12.67%	12.20%	11.54%	0.752
1	14.59%	15.30%	17.46%	
2	15.52%	14.27%	12.72%	
3–5	27.38%	27.80%	29.29%	
6–9	18.31%	29.29%	19.82%	
Charlson comorbidity index				
1	69.36%	69.51%	74.56%	0.398
2	21.28%	21.16%	15.98%	
3	6.40%	5.91%	5.92%	
≥4	2.97%	3.41%	3.55%	
Coronary heart disease	14.13%	13.41%	12.43%	0.657
Vascular disease	45.29%	43.78%	38.46%	0.067
Hypertension	41.05%	40.00%	35.50%	0.163
Nervous system disorders	14.65%	14.45%	15.68%	0.844
Stroke	0.41%	0.39%	0.30%	0.865
Diabetes	9.24%	8.35%	5.92%	0.126
Depression	10.64%	10.24%	11.54%	0.768
Anemia	0.93%	1.10%	0.30%	0.381
Platelet disorders	0.29%	0.18%	0.00%	0.533
Osteoporosis	5.64%	8.54%	8.28%	0.004
Cataract	2.50%	2.32%	3.55%	0.419

Note: *Two-sided P-values from analysis of variance for continuous variables and from the chi-squared test for categorical variables.

Abbreviations: BMI, body mass index; SD, standard deviation; GOLD, Global Initiative for Chronic Obstructive Lung Disease.

categories. The disease stage was not significantly associated with pack-years, the number of co-morbidities, Charlson's comorbidity index, and various concomitant diseases.

Tables 1B and C show that disease stage was associated significantly with almost every type of medication and resource use. Patients with more severe COPD were more likely to use various types of pulmonary maintenance medication, oxygen at home, exacerbation medication, consult their GP more often (on a scheduled and unscheduled basis), visit the emergency room with resulting admission more often, and were more likely to have been admitted for any reason to the hospital in the last year. The use of statins and other cardiovascular medication was more frequent in patients with less severe disease.

There was no difference between the dataset used for developing the algorithm and the validation dataset.

Regression results

In the final model (Table 2), the parallel regression assumption was violated for five variables (age, gender, xanthine and oxygen use, and two categories of the BMI variable, overweight and obese). In these cases, a coefficient was estimated for each dichotomization (GOLD III/IV versus GOLD II and GOLD IV versus GOLD II/III). For all other variables, one coefficient was estimated for both dichotomizations.

In the multivariable analysis, a higher risk of more severe COPD was significantly associated with being younger, being male, having a lower BMI, having quit smoking, suffering from osteoporosis, using oxygen, courses of oral steroids, having been hospitalized in the previous year, and certain types of respiratory maintenance medication (long- and short-acting bronchodilators, xanthines, leukotriene modifiers, oral steroids).

The thresholds were not statistically significantly different from 0 (*P*-values: 0.466 and 0.338, see Table 2). However, the first threshold was significantly lower than the second (−0.25075 versus 0.40289, *P* = 0.0422).

The following variables were not maintained in the final model: inhaled corticosteroids, mucolytics, cardiovascular medication, courses of antibiotics, most co-morbidities (except osteoporosis), and healthcare resource use (emergency room, GP consultations).

Predictive performance

Sixty three percent of patients in stage II, 53% of patients in stage III and 8% of patients in stage IV were correctly identified with the final model (see Table 3A). The overall agreement was 53% (unweighted) to 77% (weighted), leading to 'slight' to 'fair' kappa statistics of 0.151 (unweighted) and 0.230 (weighted).

Table 1B Medication per GOLD stage in full sample (development and validation sets combined)

Stage	II n = 1720 (47%)	III N = 1640 (44%)	IV n = 338 (9%)	P-value*
No short-acting bronchodilators	33.78%	20.91%	13.31%	0.000
1 short-acting bronchodilator	36.10%	37.56%	26.04%	
2 short-acting bronchodilators	30.12%	41.52%	60.65%	
Long-acting bronchodilator (yes/no)	62.03%	70.55%	75.74%	0.000
Inhaled corticosteroids (yes/no)	62.21%	69.51%	76.33%	0.000
Other steroids (yes/no)	5.76%	10.00%	17.16%	0.000
Xanthine (yes/no)	9.94%	19.33%	20.12%	0.000
Leukotriene modifier (yes/no)	2.79%	5.67%	8.58%	0.000
Mucolytics	4.19%	5.85%	5.33%	0.084
Home oxygen (yes/no)	1.05%	3.05%	9.17%	0.000
Statins (yes/no)	19.13%	16.52%	13.61%	0.020
Cardiovascular medication (yes/no)	51.16%	54.82%	48.82%	0.020
Courses of oral steroids				
0	70.76%	59.45%	52.66%	0.000
1	17.44%	22.68%	23.08%	
2	6.22%	9.02%	8.84%	
≥3	5.58%	8.84%	11.54%	
Courses of antibiotics				
0	53.31%	44.27%	43.20%	0.000
1	22.97%	24.94%	25.15%	
2	12.91%	16.10%	16.27%	
≥3	10.81%	14.70%	15.38%	

Note: *Two-sided *P*-values from the chi-squared test for categorical variables.

Abbreviation: GOLD, Global Initiative for Chronic Obstructive Lung Disease.

Table 1C Health care resource use per GOLD stage in full sample (development and validation sets combined)

Stage	II n = 1720 (47%)	III n = 1640 (44%)	IV n = 338 (9%)	P-value*
Scheduled GP visits				
0	20.17%	14.82%	12.43%	0.000
1	18.72%	15.85%	15.09%	
2	23.43%	23.05%	17.75%	
3–4	23.49%	28.78%	31.36%	
≥5	14.19%	17.50%	23.37%	
Unscheduled GP visits				
0	74.71%	70.12%	64.50%	0.003
1	13.14%	16.46%	18.64%	
2	6.40%	6.65%	9.17%	
≥3	5.76%	6.77%	7.69%	
Emergency room (no admission)				
0	92.40%	90.47%	89.88%	0.071
1	4.97%	6.02%	8.04%	
≥2	2.63%	3.50%	2.08%	
Emergency room and admission				
0	91.87%	88.19%	82.74%	0.000
1	6.26%	9.04%	11.90%	
≥2	1.87%	2.77%	5.36%	
Direct hospital admissions (yes/no)	3.39%	4.55%	7.14%	0.006

Note: *Two-sided P-values from the chi-squared test for categorical variables.

Abbreviations: GOLD, Global Initiative for Chronic Obstructive Lung Disease; GP, general practitioner.

Table 2 Regression results (final model)

	Parallel regression assumption holds		Parallel regression assumption violated			
	Shared by both dichotomizations		Stages III/IV versus stage II		Stage IV versus stages II/III	
	Coefficient	P	Coefficient	P	Coefficient	P
Age ²			−0.00007	0.122	−0.00055	0.000
Male			0.42304	0.000	0.84718	0.000
Underweight (BMI < 18.5)	Reference category					
Low weight (18.5 > BMI < 21)	−0.37425	0.110				
Normal BMI	−0.69927	0.001				
Overweight (25 < BMI < 30)			−0.95891	0.000	−1.77506	0.000
Obese (BMI > 30)			−1.18717	0.000	−1.91164	0.000
Current smoker	−0.15501	0.097				
Osteoporosis	0.33120	0.042				
Long-acting bronchodilator	0.26222	0.003				
No short-acting bronchodilator	Reference category					
1 short-acting bronchodilator	0.33822	0.001				
2 short-acting bronchodilators	0.85445	0.000				
Leukotriene modifier	0.66102	0.001				
Xanthine			0.61460	0.000	−0.07596	0.706
Oral steroids (maintenance)	0.30360	0.045				
No incidental courses of oral steroids	Reference category					
1 course of oral steroids	0.22220	0.037				
2 courses of oral steroids	0.29220	0.064				
≥3 courses of oral steroids	0.21776	0.184				
Oxygen			0.75863	0.000	1.59885	0.000
Any hospital admissions in previous year	0.34862	0.081				
Threshold			−0.25075	0.466	0.40289	0.338
N	2423					
Log likelihood	−2057.973					
Wald test for model significance	0.0000					

Abbreviation: BMI, body mass index.

Table 3A Predictive performance: predicted and observed GOLD stage in validation data set (PPO ordered logit model estimated in full development data set)

	Predicted GOLD stage					Unweighted	Weighted
	II	III	IV	Total			
Observed GOLD stage					Overall agreement	53%	77%
II	368	205	7	580	Kappa	0.151	0.230
	63%	35%	1%	100%	95% CI	0.103–198	0.196–0.263
III	260	283	4	547			
	46%	53%	1%	100%			
IV	28	82	10	120			
	24%	68%	8%	100%			
Total	656	570	21	1,247			

Abbreviations: CI, confidence interval; GOLD, Global Initiative for Chronic Obstructive Lung Disease.

The observed values of FEV₁%-predicted were slightly higher for COPD stage II patients who were correctly classified by the model than for patients who were incorrectly classified (59.4 versus 57.7%, respectively). In stage III, the opposite was observed (FEV₁%-predicted was 39.4% in the correctly classified versus 41.2% in the incorrectly classified), and in stage IV, FEV₁ was comparable among the correctly and incorrectly classified groups (24.9% and 25.6%).

Sensitivity analyses

In the balanced sample (regression results not presented), agreement was lower for stages II and III than in the unbalanced sample, but higher for stage IV: 53%, 33% and 57% respectively (Table 3B). The overall agreement was 48% (unweighted) to 68% (weighted). The kappa statistics were 'fair' with 0.215 and 0.290, respectively.

In the binary logit model (Table 3C), agreement for patients with moderate and severe/very severe COPD was 54% and 68%. The kappa-statistic was 'fair': 0.228.

Discussion

Our study has two important findings. Firstly, the variables that were independently related to more severe COPD defined by lung function were a lower age, male gender, a lower

BMI, being an ex-smoker, having osteoporosis, having been hospitalized in the previous year, using oxygen and certain types of respiratory maintenance medication (long- and short-acting bronchodilators, xanthines, leukotrien modifiers, oral steroids). Other variables expected to be associated with lung function impairment, such as other resource use variables, were not maintained in the final model. Secondly, the performance of the final model was such that the confidence in using the selected variables to adjust for COPD severity in the absence of lung function parameters was judged to be limited.

Of the variables in the final model, the impact of age may be partly due to a healthy survivor effect, whereas long-acting bronchodilators, multiple short-acting bronchodilators, oral steroids and xanthines are clearly indicated for treating more advanced stages of COPD. Patients with more severe disease are more likely to have quit smoking. A low BMI, often associated with loss of muscle mass is well known to be more frequent in severe COPD and the higher prevalence of osteoporosis might be a side-effect of a long history of corticosteroid use.

We used a partial proportional odds model instead of the standard ordered logit model. This made it possible to estimate different coefficients for the probability of being in stage III/IV over stage II than for the probability of being

Table 3B Predictive performance: predicted and observed GOLD stage in validation data set (PPO ordered logit model estimated in balanced sample) (sensitivity analysis)

	Predicted GOLD stage					Unweighted	Weighted
	II	III	IV	Total			
Observed GOLD stage					Overall agreement	48%	68%
II	64	34	23	121	Kappa	0.215	0.290
	53%	28%	19%	100%	95% CI	0.142–0.288	0.205–0.370
III	42	40	38	120			
	35%	33%	32%	100%			
IV	20	32	68	120			
	17%	27%	57%	100%			
Total	126	106	129	361			

Abbreviations: CI, confidence interval; GOLD, Global Initiative for Chronic Obstructive Lung Disease.

Table 3C Predictive performance: predicted and observed GOLD stage in validation data set (Binary logit model estimated in development data set)

	Predicted GOLD stage				Unweighted
	II	III/IV	Total		
Observed GOLD stage				Overall agreement	62%
II	316	264	585	Kappa	0.228
	54%	46%	100%	95% CI	0.173–0.284
III/IV	212	455	667	c-statistic	0.614
	32%	68%	100%	95% CI	0.587–0.640
Total	528	719	1,247		

Abbreviations: CI, confidence interval; GOLD, Global Initiative for Chronic Obstructive Lung Disease.

in stage IV, if this was required. The parallel regression assumption, which states that the coefficients are equal for both dichotomizations, was violated in five instances.

In the final base case model, only 53% of patients were classified in the correct GOLD stage. Especially for patients in stage IV, the predicted stage was unlikely to be correct (8% correct). Results from sensitivity analyses with a more balanced sample or a binary logit model were only slightly better.

We chose to present both unweighted and weighted kappa statistics. In cases with more than two categories, it is customary to weight the kappa-statistics in order to penalize disagreements in terms of their seriousness – ie, a higher penalty for misclassifying a patient from stage II as a stage IV patient than as a stage III patient – whereas unweighted kappa treats all disagreements equally. However, in our study all disagreements were considered serious and the weighted overall agreement might give an overly favorable impression. In the end, the unweighted and the weighted kappa statistics were quite similar. Based on all values of the kappa statistics, the agreement between predicted and observed GOLD stages can be characterized as slight to fair.

Several explanations for these findings may be considered. Firstly, the cut-off points between GOLD stages are inevitably somewhat arbitrary and artificial, especially as the decline in FEV₁ is a continuous process. A patient with a FEV₁-%-predicted of 49% (GOLD stage III) is probably less similar to a patient with a value of 31% (also GOLD stage III) than to someone with a value of 51% (GOLD stage II). If this were an important explanation for our prediction results, misclassified patients in stage IV should have markedly better lung functions – closer to stage III – than the correctly classified patients, while the opposite should be true for misclassified patients in stage II. However, the actual differences in lung function between the correctly and incorrectly classified patients in our analysis were small (1.74%-point for stage II and 0.66%-point for stage IV).

The second explanation concerns the source of our data, which was a clinical trial. It is conceivable that patients with very severe disease and many symptoms were less willing to participate in the trial. This could then obscure some of the associations between disease severity and the predictors. This problem would be expected to occur particularly when trying to predict GOLD stage IV. However, our models did not perform well at distinguishing moderate from severe patients either. In the balanced sample analysis, misclassifications occurred equally often in each GOLD stage. Furthermore, the proportion of very severe patients in the sample does not appear to be low compared to the proportion in the general population of COPD patients in Finland,³⁵ The Netherlands,³⁶ Greece,³⁷ the United Kingdom,³⁸ and a combination of European and North-American countries.³⁹ Moreover, a broad range of patients was allowed to participate. For examples, patients were permitted to use all respiratory medications concomitantly during the trial except inhaled anticholinergic drugs, thus closely resembling routine care. Altogether, this protocol makes the UPLIFT a suitable trial for this study.

The third explanation would be the far from perfect association of resource use, which is often driven by symptoms and exacerbations, with GOLD stage. Patients with a relatively good lung function do not necessarily experience fewer symptoms than patients with worse lung function. Indeed, the UPLIFT sample contains a non-negligible number of patients with very severe COPD who do use little or no maintenance medication, as well as patients with moderate disease who use four or five different types of medications. Overall, we observed an independent association of GOLD stage with several respiratory medications, but not with other types of resource use such as ER visits and physician consults in our multivariable model, which was not expected beforehand. In univariate analyses, patients with more severe COPD were more likely to have higher COPD-related resource use (almost all types of medication use, courses of oral steroids

and antibiotics, scheduled and unscheduled GP visits and hospital admissions with and without visits to the emergency room), not to use statins and other cardiovascular medication, to be younger, male, underweight, ex-smoker, and suffer from osteoporosis.

Several database studies have attempted to adjust for possible differences in COPD severity in the absence of lung function data. In an article comparing the assessment of COPD patients in the UK General Practice Research Database with the clinical opinion of the patient's GP, Soriano et al¹⁹ based their severity classification on medication use. Sin and Tu²⁰ assessed the effects of ICS use on mortality and applied medication use, ER visits and physician services as proxies of disease severity. Similarly, Suissa²¹ adjusted only for age, sex and medication use. Breekveldt-Postma et al²² identified determinants of patient's persistence with ICS therapy. They considered hospital admissions and medication use to be proxies for disease severity. In a study relating COPD severity with cardiovascular disease, Curkendall et al²³ assumed that COPD severity could be defined as the likelihood of being hospitalized, given the relationship with mortality. They concluded that this probability of hospitalization is associated with medication and oxygen use, previous hospitalizations, recent exacerbations, pneumonia and lung emphysema. Griffin et al²⁴ assessed the effects of tiotropium compared to combined ipratropium and salbutamol on exacerbations and hospitalizations. They adjusted for a combination of resource use data and risk factors. Based on the current analyses, we conclude that the variables used in these database studies cannot be relied upon to adequately adjust for COPD severity in terms of lung function.

Conclusion

Data from a well-controlled trial setting indicated that COPD severity defined by lung function thresholds cannot be reliably predicted from patient characteristics and their previous healthcare use. This limitation should be considered when interpreting findings from database studies, and additional research should explore other methods allowing accounting for COPD severity.

Acknowledgment/disclosure

LG received research grants from multiple pharmaceutical companies to perform cost-effectiveness studies. CB and KZ are employees of Pfizer Inc. BM is an employee of Boehringer Ingelheim GmbH. MR received research grants from multiple pharmaceutical companies to perform cost-effectiveness

studies of asthma/COPD medication and smoking cessation medication.

This study was funded by Boehringer Ingelheim International GmbH and Pfizer.

References

1. Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management and prevention of COPD. Available from: <http://www.guideline.gov/content.aspx?id=25648>.
2. Anthonisen NR, Wright EC, Hodgkin JE. Prognosis in chronic obstructive pulmonary disease. *Am Rev Respir Dis*. 1986;133(1):14–20.
3. Burge PS, Calverley PM, Jones PW, Spencer S, Anderson JA, Maslen TK. Randomised, double blind, placebo controlled study of fluticasone propionate in patients with moderate to severe chronic obstructive pulmonary disease: the ISOLDE trial. *BMJ*. 2000;320(7245):1297–1303.
4. Dewan NA, Rafique S, Kanwar B, et al. Acute exacerbation of COPD: Factors associated with poor treatment outcome. *Chest*. 2000;117(3):662–671.
5. Andersson F, Borg S, Jansson SA, et al. The costs of exacerbations in chronic obstructive pulmonary disease (COPD). *Respir Med*. 2002;96(9):700–708.
6. Ferrer M, Alonso J, Morera J, et al. Chronic obstructive pulmonary disease stage and health-related quality of life. the quality of life of chronic obstructive pulmonary disease study group. *Ann Intern Med*. 1997;127(12):1072–1079.
7. Jenkins C, Rodriguez-Roisin R. Quality of life, stage severity and COPD. *Eur Respir J*. 2009;33(5):953–955.
8. Stahl E, Lindberg A, Jansson SA, et al. Health-related quality of life is related to COPD disease severity. *Health Qual Life Outcomes*. 2005;3:56.
9. Ekberg-Aronsson M, Lofdahl K, Nilsson JA, Lofdahl CG, Nilsson PM. Hospital admission rates among men and women with symptoms of chronic bronchitis and airflow limitation corresponding to the GOLD stages of chronic obstructive pulmonary disease – a population-based study. *Respir Med*. 2008;102(1):109–120.
10. Ekberg-Aronsson M, Pehrsson K, Nilsson JA, Nilsson PM, Lofdahl CG. Mortality in GOLD stages of COPD and its dependence on symptoms of chronic bronchitis. *Respir Res*. 2005;6:98.
11. Antonelli-Incalzi R, Imperiale C, Bellia V, et al. Do GOLD stages of COPD severity really correspond to differences in health status? *Eur Respir J*. 2003;22(3):444–449.
12. Medinas Amoros M, Mas-Tous C, Renom-Sotorra F, Rubi-Ponseti M, Centeno-Flores MJ, Gorris-Dolz MT. Health-related quality of life is associated with COPD severity: A comparison between the GOLD staging and the BODE index. *Chron Respir Dis*. 2009;6(2):75–80.
13. Rutten-van Mölken MPMH, Oostenbrink JB, Tashkin DP, Burkhart D, Monz BU. Does quality of life of COPD patients as measured by the generic EuroQol five-dimension questionnaire differentiate between COPD severity stages? *Chest*. 2006;130(4):1117–1128.
14. 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(10):1005–1012.
15. 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(9691):704–711.
16. Jones RC, Donaldson GC, Chavannes NH, et al. Derivation and validation of a composite index of severity in chronic obstructive pulmonary disease: the DOSE Index. *Am J Respir Crit Care Med*. 2009;180(12):1189–1195.
17. Briggs A, Spencer M, Wang H, Mannino D, Sin DD. Development and validation of a prognostic index for health outcomes in chronic obstructive pulmonary disease. *Arch Intern Med*. 2008;168(1):71–79.

18. Eisner MD, Trupin L, Katz PP, et al. Development and validation of a survey-based COPD severity score. *Chest*. 2005;127(6):1890–1897.
19. Soriano JB, Maier WC, Visick G, Pride NB. Validation of general practitioner-diagnosed COPD in the UK general practice research database. *Eur J Epidemiol*. 2001;17(12):1075–1080.
20. Sin DD, Tu JV. Inhaled corticosteroids and the risk of mortality and readmission in elderly patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2001;164(4):580–584.
21. Suissa S. Effectiveness of inhaled corticosteroids in chronic obstructive pulmonary disease: Immortal time bias in observational studies. *Am J Respir Crit Care Med*. 2003;168(1):49–53.
22. Breeckveldt-Postma NS, Gerrits CM, Lammers JW, Raaijmakers JA, Herings RM. Persistence with inhaled corticosteroid therapy in daily practice. *Respir Med*. 2004;98(8):752–759.
23. Curkendall SM, Lanes S, de Luise C, et al. Chronic obstructive pulmonary disease severity and cardiovascular outcomes. *Eur J Epidemiol*. 2006;21(11):803–813.
24. Griffin J, Lee S, Caiado M, Kesten S, Price D. Comparison of tiotropium bromide and combined ipratropium/salbutamol for the treatment of COPD: A UK general practice research database 12-month follow-up study. *Prim Care Respir J*. 2008;17(2):104–110.
25. Decramer M, Celli B, Tashkin DP, et al. Clinical trial design considerations in assessing long-term functional impacts of tiotropium in COPD: The UPLIFT trial. *COPD*. 2004;1(2):303–312.
26. van Schayck CP, Halbert RJ, Norderdyke RJ, Isonaka S, Maroni J, Nonikov D. Comparison of existing symptom-based questionnaires for identifying COPD in the general practice setting. *Respirology*. 2005;10(3):323–333.
27. Tinkelman DG, Price DB, Norderdyke RJ, et al. Symptom-based questionnaire for differentiating COPD and asthma. *Respiration*. 2006;73(3):296–305.
28. Price DB, Tinkelman DG, Norderdyke RJ, Isonaka S, Halbert RJ; COPD Questionnaire Study Group. Scoring system and clinical application of COPD diagnostic questionnaires. *Chest*. 2006;129(6):1531–1539.
29. Martinez FJ, Foster G, Curtis JL, et al. Predictors of mortality in patients with emphysema and severe airflow obstruction. *Am J Respir Crit Care Med*. 2006;173(12):1326–1334.
30. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373–383.
31. Long JS. *Regression Models for Categorical and Limited Dependent Variables (Advanced Quantitative Techniques in the Social Sciences)*. Thousand Oaks: Sage; 1997.
32. Williams R. Generalized Ordered logit/partial proportional odds models for ordinal dependent variables. *The Stata Journal*. 2006;6(1):58–82.
33. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. 3rd ed. New York: John Wiley & Sons; 2003.
34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
35. Vasankari TM, Impivaara O, Heliovaara M, et al. No increase in the prevalence of COPD in two decades. *Eur Respir J*. 2010;36(4):766–773.
36. Hoogendoorn M, Feenstra TL, Schermer TR, Hesselink AE, Rutten-van Molken MP. Severity distribution of chronic obstructive pulmonary disease (COPD) in Dutch general practice. *Respir Med*. 2006;100(1):83–86.
37. Minas M, Hatzoglou C, Karetsi E, et al. COPD prevalence and the differences between newly and previously diagnosed COPD patients in a spirometry program. *Prim Care Respir J*. 2010;19(4):363–370.
38. Shahab L, Jarvis MJ, Britton J, West R. Prevalence, diagnosis and relation to tobacco dependence of chronic obstructive pulmonary disease in a nationally representative population sample. *Thorax*. 2006;61(12):1043–1047.
39. Agusti A, Calverley PM, Celli B, et al. Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respir Res*. 2010;11:122.

International Journal of COPD

Publish your work in this journal

The International Journal of COPD is an international, peer-reviewed journal of therapeutics and pharmacology focusing on concise rapid reporting of clinical studies and reviews in COPD. Special focus is given to the pathophysiological processes underlying the disease, intervention programs, patient focused education, and self management protocols.

Submit your manuscript here: <http://www.dovepress.com/international-journal-of-copd-journal>

Dovepress

This journal is indexed on PubMed Central, MedLine and CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.