

Prediction models for exacerbations in different COPD patient populations: comparing results of five large data sources

Martine Hoogendoorn¹
 Talitha L Feenstra^{2,3}
 Melinde Boland¹
 Andrew H Briggs⁴
 Sixten Borg⁵
 Sven-Arne Jansson⁶
 Nancy A Risebrough⁷
 Julia F Slejko⁸
 Maureen PMH Rutten-van
 Mólken¹

¹Institute for Medical Technology Assessment (iMTA)/Erasmus School of Health Policy & Management (ESHPM), Erasmus University Rotterdam, Rotterdam, the Netherlands; ²Department for Prevention and Health Services Research, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands; ³Department of Epidemiology, Groningen University, University Medical Centre Groningen, Groningen, the Netherlands; ⁴Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK; ⁵Health Economics Unit, Department of Clinical Sciences Malmö, Lund University, Lund, Sweden; ⁶Department of Public Health and Clinical Medicine, Occupational and Environmental Medicine, The OLIN Unit, Umeå University, Umeå, Sweden; ⁷ICON Health Economics, Toronto, Canada; ⁸Department of Pharmaceutical Health Services Research, University of Maryland School of Pharmacy, Baltimore, MD, USA

Correspondence: Martine Hoogendoorn
 Institute for Medical Technology Assessment
 (iMTA)/Erasmus School of Health Policy
 & Management (ESHPM), Erasmus
 University Rotterdam, PO Box 1738, 3000
 DR Rotterdam, the Netherlands
 Tel +31 10 408 8871
 Email hoogendoorn@imta.eur.nl

Background and objectives: Exacerbations are important outcomes in COPD both from a clinical and an economic perspective. Most studies investigating predictors of exacerbations were performed in COPD patients participating in pharmacological clinical trials who usually have moderate to severe airflow obstruction. This study was aimed to investigate whether predictors of COPD exacerbations depend on the COPD population studied.

Methods: A network of COPD health economic modelers used data from five COPD data sources – two population-based studies (COPDGene[®] and The Obstructive Lung Disease in Norrbotten), one primary care study (RECODE), and two studies in secondary care (Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoint and UPLIFT) – to estimate and validate several prediction models for total and severe exacerbations (= hospitalization). The models differed in terms of predictors (depending on availability) and type of model.

Results: FEV₁% predicted and previous exacerbations were significant predictors of total exacerbations in all five data sources. Disease-specific quality of life and gender were predictors in four out of four and three out of five data sources, respectively. Age was significant only in the two studies including secondary care patients. Other significant predictors of total exacerbations available in one database were: presence of cough and wheeze, pack-years, 6-min walking distance, inhaled corticosteroid use, and oxygen saturation. Predictors of severe exacerbations were in general the same as for total exacerbations, but in addition low body mass index, cardiovascular disease, and emphysema were significant predictors of hospitalization for an exacerbation in secondary care patients.

Conclusions: FEV₁% predicted, previous exacerbations, and disease-specific quality of life were predictors of exacerbations in patients regardless of their COPD severity, while age, low body mass index, cardiovascular disease, and emphysema seem to be predictors in secondary care patients only.

Keywords: COPD, exacerbations, modeling, hospitalizations, validation

Introduction

It is well known that the progression of COPD may be accompanied by exacerbations, that is, an acute worsening of symptoms. Exacerbations are associated with an accelerated decline in lung function,^{1,2} increase in mortality,^{3,4} significant impairment of health-related quality of life,⁵⁻⁷ and increased health care utilization and associated costs.⁸⁻¹⁰ Consequently, reducing exacerbations is one of the most important treatment goals in COPD from both a clinical and an economic perspective.^{11,12} Because not all patients experience exacerbations, identification of patients who are at high risk of exacerbations is important to use treatment options in an efficient way.

Health economic decision models for COPD are used for evaluating long-term effectiveness and cost-effectiveness of treatment options for COPD. Accurate and precise estimation of the exacerbation risk in these models is very important because exacerbations are associated with high health care costs and therefore strongly influence the cost outcomes of the models. However, most of the currently available cost-effectiveness models include an exacerbation risk specified by degree of airflow obstruction¹³ only and they are not able to distinguish high-risk patients based on other relevant predictors, such as previous exacerbations.¹⁴ Therefore, the focus of the Fourth Annual Meeting of the International COPD Health Economic Modelling Network in 2015 was to develop prediction models for exacerbations including several relevant patient and disease characteristics as predictors.

Most of the previously published studies investigating predictors of exacerbations were done in COPD patients participating in pharmacological clinical trials who usually have moderate to severe airflow obstruction.^{14–19} A few studies were done in primary care patients with less severe COPD.^{20–22} Because the studies included different candidate predictors or used different definitions for predictors, it is difficult to conclude whether predictors of exacerbations differ between patient populations with varying disease severity.

The aim of the current study was to estimate prediction models for the total number of exacerbations and severe exacerbations using five large sources of patient-level data and compare the estimated models between patient populations.

Methods Procedure

Since 2011, a worldwide network of people involved in health economic modeling for COPD (ie, COPD modeling teams, employees of pharmaceutical companies interested in COPD modeling, clinicians, health economists, and epidemiologists) gathered together for three one-day meetings with the aim to discuss and compare the different available COPD models and share best practices about COPD modeling.^{13,23} In May 2015 participants in this COPD modeling network were contacted to explore their interest in participating in a modeling exercise for the fourth COPD modeling meeting. To participate in this so-called modeling challenge, participants needed to have access to a database with patient-level data with the following characteristics: 1) a minimum of about 500 patients, 2) follow-up of at least 1 year, 3) moderate and severe exacerbations measured, and 4) several demographic and clinical patient characteristics available. For the first part of the challenge, participants were asked

to estimate several pre-specified prediction models including one or multiple predictors and using different statistical methods. In the second part of the challenge, the estimated prediction models were validated. A structured Microsoft Excel file was used to collect the results in a uniform way before the meeting. During the meeting the results were presented and discussed.

Sources of data

Data from five different data sources were used for the modeling challenge: two population-based studies – the COPDGene[®] study and the Obstructive Lung Disease in Norrbotten (OLIN) study – one study in primary care patients – the RECODE trial – and two studies including secondary care patients – the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) study and the UPLIFT trial.^{24–28} The COPDGene study is a multi-institutional study of past and current smokers to identify the genetic factors that control the development and progression of COPD. The primary analysis cohort dataset consists of about 10,300 smokers in the general population. For this modeling challenge a subgroup of 3,756 patients with a diagnosis of COPD were included, who were on average followed for 4.7 years.²⁴ The OLIN study is a population-based screening study for COPD in the general population in northern Sweden. The prevalence of more severe patients is therefore low, reflective of actual prevalence. For the current study, the data of patients with a COPD diagnosis enrolled in 2005 and their follow-up data for 2006 were used. Observations of severe exacerbations were scarce, and therefore no separate analysis for severe exacerbations was performed.²⁵ The RECODE trial was a 2-year cluster-randomized trial in which 20 primary care teams were randomized to an intervention group of general practitioner practices that implemented an integrated care program for COPD and 20 teams were randomized to a usual care group.²⁶ The ECLIPSE study was a non-interventional, longitudinal prospective 3-year study in COPD patients aged 40–75 years with a baseline post-bronchodilator forced expiratory volume in 1 s (FEV₁)% predicted <80%, baseline post-bronchodilator FEV₁/forced vital capacity (FVC) <0.7, and a smoking history of at least 10 pack-years.²⁷ The UPLIFT trial was a 4-year randomized controlled trial comparing tiotropium versus placebo in patients with a diagnosis of COPD (FEV₁/FVC <70%), age ≥40 years, >10 pack-years, and an FEV₁% predicted <70%.²⁸ The study protocols of all five studies were approved by the relevant ethics and review boards of the participating centers.

Exacerbations

Participants were asked to use the following definition for an exacerbation: a moderate exacerbation was defined as an increase in symptoms requiring a visit to a health care provider and a course of antibiotics and/or oral steroids. A severe exacerbation was defined as an exacerbation requiring hospitalization.

Predictions models

Prediction models were estimated separately for total and severe exacerbations. For each data source, a randomly selected 67% of the patient population was used to estimate the prediction models, while the remaining 33% was set aside for validation purposes.

Five different prediction models for the annual exacerbation rate were estimated using negative binomial regression²⁹ with the natural logarithm of the total time at risk in years as offset variable and the number of exacerbations as outcome. The time at risk was defined as total time in the study, which was shorter than the duration of the study for patients who dropped out or died before the end of the study. The estimated regression coefficients for the predictors were transformed into incidence rate ratio (IRR) by taking the exponent of the estimated coefficient. All variables included as predictors were measured at baseline. The five different prediction models varied in the number of predictors and the type of model used. The first two prediction models were pre-specified and included the same predictors for all data sources. For the first model, groups were asked to estimate a model including the predictors FEV₁% predicted at baseline and treatment for the two trials, that is, RECODE and UPLIFT. For the second model, groups were asked to include the same patient and clinical parameters as in a previously published prediction model by Briggs et al,¹⁶ that is, sex, age, FEV₁% predicted, total number of (severe) exacerbations in the year prior to baseline (depending on the outcome used), disease-specific quality of life, body mass index (BMI) <20 kg/m², history of cardiovascular disease and treatment, if applicable. For model 3, groups were asked to include all patient and clinical parameters from model 2 plus other possible relevant variables available in the data source to be chosen by the modeling team, for example, dyspnea, COPD duration, and other comorbidities. The latter predictors were different for the different data sources. In the fourth model, the impact of using a different statistical method was explored. Model 4 included the same predictors as model 3 but used a zero-inflated binomial regression model with an indicator of previous exacerbations at baseline in

the zero-model. With model 5, the impact of two different definitions for previous exacerbations was assessed using multilevel negative binomial regression. Model 5 included the same predictors as model 3 and had two variants: one using exacerbations prior to baseline as predictor (model 5A) and one using exacerbations in the previous period as predictor (model 5B; using the exacerbation rate in Year 1 to predict the rate in Year 2, the rate in Year 2 to predict the rate in Year 3, etc.).

Validation

For the validation part, participants were asked to use the remaining 33% of the population in their data source. First the mean observed exacerbation rate was calculated using the observation time of each patient as a weight, that is, patients with a longer follow-up have a higher weight than patients with a short follow-up. In addition, the predicted exacerbation rate for each patient based on each of the five different models was calculated by filling in the estimated regression equations for each individual patient. Thereafter, the mean predicted exacerbation rate over all patients was determined. Finally, absolute errors were calculated as the absolute difference between the individual observed and predicted exacerbation rates for each patient. The mean absolute error (MAE) between the observed and predicted rates was calculated using the observation time of the individual patient as a weight. If the MAE is small, the observed and predicted exacerbation rates are fairly similar. If the MAE is large, the predicted rate for patients is substantially different from the observed rate.

Results

Characteristics of the patients included in the five different participating data sources are shown in Table 1. For the ECLIPSE and UPLIFT studies, the mean FEV₁% predicted was about 48%, and more than half of the patients had severe to very severe airflow obstruction. Patients in the population-based OLIN study had the highest mean FEV₁% predicted (76%). In this data source, only 7% of the patients were classified as having severe or very severe airflow obstruction. Mean FEV₁% predicted of the population-based COPD Gene study and the primary care-based RECODE study were 57% and 68%, respectively.

Figure 1 shows the percentage of patients with at least one exacerbation during follow-up. The lowest percentages were found for the OLIN study with a follow-up duration of 1 year. Percentages in ECLIPSE were highest, although UPLIFT had a longer follow-up.

Table 1 Baseline characteristics[#] of the patients in the five data sources, data are mean or %

	COPDGene ²⁴	OLIN ²⁵	RECODE ²⁶	ECLIPSE ²⁷	UPLIFT ²⁸
N	3,756	449	1,086	2,164	5,799
Male (%)	44	60	46	65	75
Age	64	63	68	63	64
Post FEV ₁ % predicted	57.3	75.8	67.8	48.3	47.6
GOLD stages based on FEV ₁ %					
Mild	16	41	24	0	0
Moderate	41	53	53	44	46
Severe	26	6	19	42	45
Very severe	12	1	3	14	9
Smoker (%)	37	Na	37	36	30
BMI <20 (%)	5	3	Na	11	11
History of cardiovascular disease (%)	9.8	25	16	33	52
SGRQ total score	36	Na	36	50	46
Exacerbations in the year prior to baseline					
Total exacerbations	0.64	0.30	0.37	1.21	0.85
Severe exacerbations	0.18	0.02	0.02*	0.22	0.25

Notes: [#]Only characteristics that were available in all or almost all databases were included in this table. *In 3 months.

Abbreviations: BMI, body mass index; ECLIPSE, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints; FEV₁, forced expiratory volume in 1 second; GOLD, Global Initiative for chronic Obstructive Lung Disease; Na, not available; OLIN, Obstructive Lung Disease in Norrbotten; SGRQ, St George's Respiratory Questionnaire.

Figure 2 shows the mean exacerbation rates during follow-up for all five data sources. The number of exacerbations per patient-year ranged from 0.18 in the OLIN study to 1.20 in the ECLIPSE study. The number of severe exacerbations per patient-year varied between 0.02 for OLIN and 0.26 for ECLIPSE.

Results of prediction model 1 including only FEV₁% predicted and treatment, if applicable, showed that for all five data sources FEV₁% was a significant predictor of total and severe exacerbations. Model 2, that is, with addition of sex, age, exacerbations prior to baseline, disease-specific quality of life, low BMI, and history of cardiovascular disease showed that besides FEV₁% predicted, exacerbations prior to baseline and disease-specific quality of life were important predictors of exacerbations. Results of model 3, which also allowed inclusion of

other data source-specific variables besides the fixed set of predictors of model 2, are shown in Table 2 for total exacerbations and Table 3 for severe exacerbations. Both tables show IRRs, while more detailed information on coefficients, standard errors, and *p*-values for the different predictors are presented in the Supplementary materials (Tables S1–S5).

FEV₁% predicted and exacerbations prior to baseline were the most important significant predictors of total exacerbations in all five data sources (*p*-values <0.001) (Table 2). Disease-specific quality of life, that is, St George's Respiratory Questionnaire (SGRQ) total score was also an important significant predictor in all four databases for which it was available. For the other significant predictors, there was more variation in *p*-values. Sex was a predictor in three out of five databases with female patients having higher

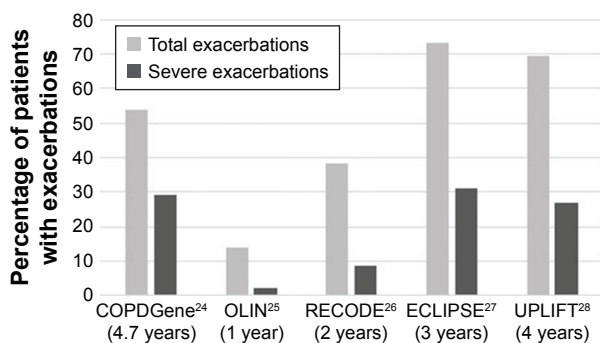


Figure 1 Percentage of patients with at least one (severe) exacerbation during follow-up with the duration of follow-up presented in brackets.

Abbreviations: ECLIPSE, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints; OLIN, Obstructive Lung Disease in Norrbotten.

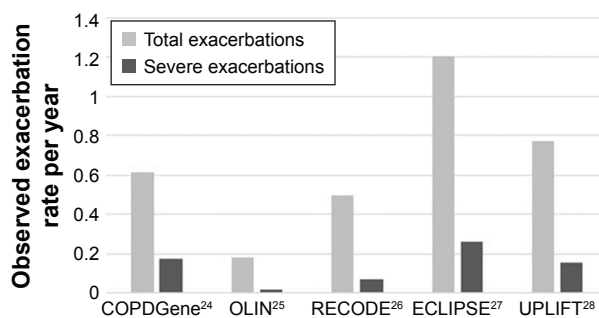


Figure 2 Mean annual total and severe exacerbation rates during follow-up. (Rates are calculated as the sum of exacerbations over all patients divided by the sum of follow-up time to correct for patients with a short follow-up time.)

Abbreviations: ECLIPSE, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints; OLIN, Obstructive Lung Disease in Norrbotten.

Table 2 Prediction models for total exacerbations including a fixed set of predictors and other database-specific predictors (results from multivariate analysis)

	IRRs [#]				
	COPDGene ²⁴	OLIN ²⁵	RECODE ²⁶	ECLIPSE ²⁷	UPLIFT ²⁸
Fixed set of predictors					
Sex (1= female)	1.23**	2.86*	1.20	1.31***	1.09
Age (years)	1.00	1.00	1.00	1.01*	1.01***
FEV ₁ % predicted (in %)	0.99***	0.96**	0.98***	0.99***	0.99***
Number of exacerbations prior to baseline	1.75***	1.49***	1.58***	1.38***	1.25***
BMI <20 kg/m ² (1= yes)	0.98	1.65	–	1.09	1.20**
History of cardiovascular disease (1= yes)	1.07	1.13	1.06	0.98	0.99
SGRQ total score at baseline (in points)	1.01***	–	1.02**	1.01***	1.01***
Treatment group in trial (1= yes)	–	–	1.08	–	0.83***
Other database-specific predictors					
Smoker (1= yes, 0= former)	0.81**	–	1.07	0.88*	1.05
Pack-years	–	–	–	–	1.002**
Time since diagnosis (years)	–	–	–	–	1.005
Diagnosis of emphysema (1= yes)	–	–	–	–	0.97
Cough (1= yes)	1.16*	–	–	–	–
Wheeze (1= yes)	1.37***	–	–	–	–
MRC dyspnea	–	–	0.98	1.03	–
Charlson comorbidity index	–	–	0.98	–	0.98
Other comorbidities (1= yes)	–	–	–	1.00	–
6-min walking test (m)	–	–	–	1.001***	–
Physical activity IPAQ (1= low)	–	–	0.81	–	–
ICS at baseline (1= yes)	–	–	–	–	1.30***
Resting O ₂ saturation (in %)	0.97**	–	–	–	–
Fibrinogen (mg/dL)	–	–	–	1.00	–

Notes: [#]Ratios above 1 indicating an increased risk and ratios below 1 indicating a reduced risk for exacerbations. **p*<0.05, ***p*<0.01, ****p*<0.001.

Abbreviations: BMI, body mass index; ECLIPSE, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints; FEV₁, forced expiratory volume in 1 second; ICS, inhaled corticosteroid; IPAQ, International Physical Activity Questionnaire; IRR, incidence rate ratio; MRC, Medical Research Council; OLIN, Obstructive Lung Disease in Norrbotten; SGRQ, St George's Respiratory Questionnaire.

rates. Age was significant only in the two studies including secondary care patients, ECLIPSE and UPLIFT. Results for current smoking in two studies showed that current smokers had a lower incidence rate for exacerbations than former smokers, while two other studies found a non-significant higher rate for current smokers. Other predictors found to be predictive of a higher exacerbation rate available in a single data source were: increasing number of pack-years, presence of cough and wheeze, 6-min walking test, and inhaled corticosteroid (ICS) use at baseline. An increase in resting oxygen was found to be associated with a decrease in exacerbation rate in COPDGene.

Predictors of the number of severe exacerbations were in general the same as of the total number of exacerbations, but in addition low BMI, history of cardiovascular disease, presence of emphysema, and current smoking were found to be predictors in secondary care COPD patients (ECLIPSE and/or UPLIFT). Contrary to the total number of exacerbations, sex was a significant predictor of the number of severe exacerbations in only one database (COPDGene).

Results for model 4 using a zero-inflated negative binomial regression model instead of regular negative binomial regression showed that using this type of model is relevant only when the proportion of patients without any exacerbations is very high. Only for the OLIN study in which 86% of patients had no exacerbations during follow-up, the coefficient for the zero-inflated model parameter was significant. Using exacerbations in the previous period as predictor instead of using exacerbations in the year prior to baseline (model 5) did not seem to improve the model fit much. Only for UPLIFT, the study with the longest follow-up period, a slight improvement in the model fit was observed.

Validation results for the different models are shown in Table 4. In general, for total number of exacerbations the mean annual predicted rates were somewhat higher than the mean observed rates, but the MAEs of the models were large. Compared to the model including FEV₁ only (model 1), the MAEs for the models including more patient characteristics tended to be slightly lower indicating that the models with more patient characteristics resulted in slightly better

Table 3 Prediction models for severe exacerbations including a fixed set of predictors and other database-specific predictors (results from multivariate analysis)

	IRRs [#]			
	COPDGene ²⁴	RECODE ²⁶	ECLIPSE ²⁷	UPLIFT ²⁸
Fixed set of predictors				
Sex (1= female)	1.22*	0.64	1.05	1.02
Age (years)	1.00	1.02	1.03**	1.03***
FEV ₁ predicted (in %)	0.99***	0.98*	0.98***	0.97***
Number of severe exacerbations prior to baseline	2.15***	5.19*	1.99***	1.68***
BMI <20 (1= yes)	1.30	–	1.44*	1.83***
History of cardiovascular disease (1= yes)	1.26	0.64	1.38**	1.11
SGRQ total score at baseline (in points)	1.02***	1.03**	1.02***	1.01***
Treatment group in trial (1= yes)	–	0.96	–	0.86*
Other database-specific predictors				
Smoker (1= yes, 0= former)	0.99	1.23	1.11	1.23*
Pack-years	–	–	–	1.00
Time since diagnosis (years)	–	–	–	1.01
Diagnosis of emphysema (1= yes)	–	–	–	1.18*
Cough (1= yes)	1.01	–	–	–
Wheeze (1= yes)	1.34**	–	–	–
MRC dyspnea	–	1.14	0.93	–
Charlson comorbidity index	–	1.03	–	1.03
Other comorbidities (1= yes)	–	–	0.87	–
6-min walking test (meters)	–	–	1.00	–
Physical activity IPAQ (1= low)	–	0.98	–	–
ICS at baseline (1= yes)	–	–	–	1.23**
Resting O ₂ saturation	1.00	–	–	–
Fibrinogen (mg/dL)	–	–	1.00	–

Notes: [#]Ratios above 1 indicating an increased risk and ratios below 1 indicating a reduced risk for exacerbations. **p*<0.05. ***p*<0.01. ****p*<0.001.

Abbreviations: BMI, body mass index; ECLIPSE, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints; FEV₁, forced expiratory volume in 1 second; ICS, inhaled corticosteroid; IPAQ, International Physical Activity Questionnaire; IRR, incidence rate ratio; MRC, Medical Research Council; SGRQ, St George's Respiratory Questionnaire.

predictions on the individual level. For severe exacerbations, the mean annual predicted rates were higher than the mean observed rates for all models in all databases. Contrary to the results for total exacerbations, the MAEs did not seem to decrease when more patient and clinical parameters were added to the models.

Discussion

This study aimed to estimate prediction models for total and severe exacerbations using different COPD patient populations to explore whether predictors of exacerbations are different in patient populations with a different disease severity. Results showed that FEV₁ predicted and previous exacerbations were significant predictors in all data sources regardless of the severity of the COPD population. The importance of previous exacerbations as predictor of future exacerbations was already well known from studies mainly performed in secondary care patients with severe COPD.^{14–17,19} The current study showed that the number of previous exacerbations is also a good predictor in patients with less severe airflow obstruction. Even if more than 40%

of the patients were classified as having mild obstruction (OLIN studies), previous exacerbations were found to be a strong predictor. The SGRQ total score was also found to be an important significant predictor of exacerbations in four out of four data sources, which was also in line with three previously performed studies in severe COPD patients.^{14,16,17} One study in COPD patients with less severe airflow obstruction reported an association between exacerbation risk and responses to the clinical COPD questionnaire.³⁰ Previous studies already showed that patients with frequent exacerbations have a lower health-related quality of life, but in the current study health status was found to be an independent predictor of new exacerbations after adjusting for previous exacerbations. Additional analyses using the UPLIFT and RECODE data showed that especially the SGRQ symptom sub-score and the SGRQ impact sub-score were predictive of future exacerbation risk, while the SGRQ activity sub-score was not found to be a significant predictor.

Female gender was found to be a predictor of total exacerbations in three out of five data sources. In the UPLIFT trial, female gender was borderline significant (*p*=0.07).

Table 4 Model validation results for total and severe exacerbations: weighted mean observed annual exacerbation rate, mean predicted annual exacerbation rates, and weighted MAEs

	COPDGene ²⁴		OLIN ²⁵		RECODE ²⁶		ECLIPSE ²⁷		UPLIFT ²⁸	
	Rate	MAE	Rate	MAE	Rate	MAE	Rate	MAE	Rate	MAE
Total exacerbations										
Weighted mean observed rate	0.61	–	0.15	–	0.51	–	1.22	–	0.79	–
Predicted rates										
Model 1: FEV ₁ only	0.65	0.63	0.18	0.28	0.51	0.62	1.25	1.05	0.83	0.64
Model 2: fixed set of predictors	0.65	0.58	0.25	0.28	0.57	0.56	1.31	0.80	0.89	0.62
Model 3: model 2 plus data source-specific predictors	0.66	0.58	Na	Na	0.57	0.56	1.28	0.81	0.89	0.62
Model 4: zero-inflated model	0.64	0.58	0.20	0.22	0.66	0.60	1.24	0.77	0.91	0.62
Model 5A: exacerbations in the year prior to baseline	Na	Na	Na	Na	0.60	0.64	1.09	0.94	0.82	0.84
Model 5B: exacerbations in the previous period	Na	Na	Na	Na	0.56	0.61	1.23	0.95	0.83	0.82
Severe exacerbations										
Weighted mean observed rate	0.17	–	0.02	–	0.07	–	0.27	–	0.15	–
Predicted rates										
Model 1: FEV ₁ only	0.22	0.25	Na	Na	0.07	0.13	0.28	0.34	0.18	0.22
Model 2: fixed set of predictors	0.22	0.22	Na	Na	0.17	0.19	0.34	0.44	0.40	0.34
Model 3: model 2 plus data source-specific predictors	0.22	0.22	Na	Na	0.18	0.21	0.33	0.44	0.41	0.35
Model 4: zero-inflated model	0.21	0.22	Na	Na	0.24	0.23	0.30	0.41	0.34	0.30
Model 5A: exacerbations in the year prior to baseline	Na	Na	Na	Na	0.14	0.17	0.19	0.34	0.27	0.34
Model 5B: exacerbations in the previous period	Na	Na	–	–	0.09	0.13	0.19	0.35	0.24	0.32

Abbreviations: ECLIPSE, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints; FEV₁, forced expiratory volume in 1 second; MAEs, mean absolute errors; Na, not available; MRC, Medical Research Council; OLIN, Obstructive Lung Disease in Norrbotten.

For severe exacerbations, only one data source found a significant impact of gender. Seven out of nine studies found in the literature confirmed the finding that female patients have a higher number of total exacerbations.^{15–17,19,21,22,31} Little is known about the reason why women seem to have more exacerbations. It may be related to physiological or treatment-related factors,³² but it might also be explained by the health care-based definition used to identify exacerbations and the fact that women tend to seek health care more often.³³ In the current study, age was a predictor only in the two studies including patients with severe airflow obstruction. This finding seems to be in line with the literature that showed an association between age and exacerbations in five out of seven studies including patients with severe airflow obstruction,^{14–19,31} while this association was found in only one out of five studies including less severe patients.^{20–22,30,34} In none of the currently explored data sources, history of cardiovascular disease was a predictor of total exacerbations, while several studies in patients with less severe airflow obstruction did find an association.^{20–22,34} Predictors of severe exacerbations were in general the same as of total exacerbations. Due to the definition used, predictors of severe

exacerbations actually need to be interpreted as predictors of hospitalization for an exacerbation. The results showed that patients with mainly severe airflow obstruction and additional risk factors, that is, history of cardiovascular disease, low BMI, current smoking, and presence of emphysema, are more likely to be hospitalized, which is as expected.

Model validation results showed that the models predicted exacerbation rates quite well on an average level, because the mean predicted rates were comparable although somewhat higher than the mean observed rate. However, the MAEs of the models were large indicating that the models were predicting less well on an individual level. MAEs seem to decrease from model 1 to model 3, showing that adding more patient and clinical parameters seemed to improve predictions on the individual level slightly.

Based on the results of the prediction models, it is difficult to come up with interventions or treatment advice to reduce the risk of an exacerbation. Some factors, such as gender and age, cannot be influenced, while others such as FEV₁ and SGRQ total score can be influenced by treatment. They are well recognized as important treatment goals, but it is a challenge to change them substantially within a short time period.

Therefore, results of the current study mainly create awareness about which type of patients are more likely to experience exacerbations and hence should be closely monitored. It further stresses on the need for new, more targeted predictors of exacerbations like biomarkers, which were unfortunately not present in the available databases.

A strength of the current study is that exactly the same methods were used in five different data sources. By using the same type of model with a partially fixed set of predictors and ensuring variables were defined in the same way in all data sources, the comparability of the results between the different data sources, and hence populations, was improved as much as possible.

A limitation of the study is that there are slight differences in the way exacerbations are defined in the five available databases. All databases included treatment with antibiotics or oral corticosteroids in the definition. But in some databases patients needed to have an increase in respiratory symptoms lasting for a pre-specified number of days, while in other databases an unscheduled visit to a health care provider was a requirement. Next to that, the studies were performed in different countries with differences in treatment patterns and access to health care. These differences in the definitions and health care settings mainly affected the observed exacerbation rates and most likely not the analyses of predictors. Heterogeneity in the definition of a severe exacerbation might have had more impact on the results. Because the different studies were performed in different health care settings, the likelihood to be hospitalized for an exacerbation may also vary substantially between patients. This may partly explain why the predictors of total and severe exacerbations are very similar. Despite this, we found some new predictors of severe exacerbations especially in secondary care patients.

Although a large number of different predictors of exacerbations were included in the different regression models, not all potential relevant predictors were included in the analyses. Especially, information on biomarkers is lacking in the current study. This was mainly because the majority of studies were performed almost 10 years ago, when data on biomarkers were not yet collected as often as they are nowadays. Using older data might have had an impact on the absolute exacerbation rates observed in the different databases. Because several new treatment options became available in the last decade, exacerbation rates in more recent trials might be lower. However, a lower rate is unlikely to greatly influence the variables that were found to be predictors.

In conclusion, FEV₁% predicted, previous exacerbations, and disease-specific quality of life were identified

as predictors of the total number of exacerbations in COPD patients regardless of their COPD severity. In secondary care patients age was found to be a predictor of total exacerbations, and low BMI, history of cardiovascular disease, and presence of emphysema were predictors of hospitalization for an exacerbation.

Acknowledgments

The COPDGene project is supported by award numbers R01 HL089897, R01 HL089856, and K01 HL125858 from the National Heart, Lung, and Blood Institute. The ECLIPSE study was supported by GlaxoSmithKline (SCO104960/NCT00292552). Financial support for the OLIN study was received mainly from The Swedish Heart & Lung Foundation (20050428, 20090244, and 20150488), The Swedish Research Council (80586701), ALF (216371) a regional agreement between Umeå University and Norrbotten County Council (NLL-574941), Norrbotten County Council, the Swedish Asthma-Allergy Foundation, and Visare Norr. The RECODE study has been funded by a Dutch Healthcare insurance company (Stichting Achmea Gezondheidszorg) and the Netherlands Organisation for Health Research and Development (Zon-MW) (project number 171002203). The UPLIFT trial was funded by Boehringer Ingelheim. The current study was financially supported by Boehringer Ingelheim International, GlaxoSmithKline, the Netherlands, and Novartis International.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. *Thorax*. 2002;57(10):847–852.
2. Celli BR, Thomas NE, Anderson JA, et al. Effect of pharmacotherapy on rate of decline of lung function in chronic obstructive pulmonary disease: results from the TORCH study. *Am J Respir Crit Care Med*. 2008;178(4):332–338.
3. Soler-Cataluna 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(11):925–931.
4. Hoogendoorn M, Hoogenveen RT, Rutten-van Molken MP, Vestbo J, Feenstra TL. Case-fatality of COPD exacerbations: a meta-analysis and statistical modeling approach. *Eur Respir J*. 2011;37(3):508–515.
5. Miravittles M, Ferrer M, Pont A, Zalacain R, et al. Effect of exacerbations on quality of life in patients with chronic obstructive pulmonary disease: a 2 year follow up study. *Thorax*. 2004;59(5):387–395.
6. Seemungal TA, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 1998;157(5):1418–1422.
7. Spencer S, Calverley PM, Burge PS, Jones PW. Impact of preventing exacerbations on deterioration of health status in COPD. *Eur Respir J*. 2004;23(5):698–702.

8. 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.
9. Oostenbrink JB, Rutten-van Molken MP. Resource use and risk factors in high-cost exacerbations of COPD. *Respir Med.* 2004;98(9): 883–891.
10. O'Reilly JF, Williams AE, Rice L. Health status impairment and costs associated with COPD exacerbation managed in hospital. *Int J Clin Pract.* 2007;61(7):1112–1120.
11. Global Initiative for chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. 2013.
12. Criner GJ, Bourbeau J, Diekemper RL, et al. Prevention of acute exacerbations of COPD: American College of Chest Physicians and Canadian Thoracic Society Guideline. *Chest.* 2015;147(4):894–942.
13. Hoogendoorn M, Feenstra TL, Asukai Y, et al. Patient heterogeneity in health economic decision models for chronic obstructive pulmonary disease: are current models suitable to evaluate personalized medicine? *Value Health.* 2016;19(6):800–810.
14. Hurst JR, Vestbo J, Anzueto A, et al. Evaluation of COPD longitudinally to identify predictive surrogate endpoints (ECLIPSE) Investigators. Susceptibility to exacerbation in chronic obstructive pulmonary disease. *N Engl J Med.* 2010;363(12):1128–1138.
15. Beeh KM, Glaab T, Stowasser S, et al. Characterisation of exacerbation risk and exacerbator phenotypes in the POET-COPD trial. *Respir Res.* 2013;14(1):116.
16. 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.
17. Make BJ, Eriksson G, Calverley PM, et al. A score to predict short-term risk of COPD exacerbations (SCOPEX). *Int J Chron Obstruct Pulmon Dis.* 2015;10:201–209.
18. Niewoehner DE, Likhnygina Y, Rice K, et al. Risk indexes for exacerbations and hospitalizations due to COPD. *Chest.* 2007;131(1): 20–28.
19. Jenkins CR, Celli B, Anderson JA, et al. Seasonality and determinants of moderate and severe COPD exacerbations in the TORCH study. *Eur Respir J.* 2012;39(1):38–45.
20. Bertens LC, Reitsma JB, Moons KG, et al. Development and validation of a model to predict the risk of exacerbations in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis.* 2013;8:493–499.
21. Mullerova H, Shukla A, Hawkins A, Quint J. Risk factors for acute exacerbations of COPD in a primary care population: a retrospective observational cohort study. *BMJ Open.* 2014;4(12):e006171.
22. McGarvey L, Lee AJ, Roberts J, Gruffydd-Jones K, McKnight E, Haughney J. Characterisation of the frequent exacerbator phenotype in COPD patients in a large UK primary care population. *Respir Med.* 2015;109(2):228–237.
23. Hoogendoorn M, Feenstra TL, Asukai Y, et al. Cost-effectiveness models for chronic obstructive pulmonary disease: cross-model comparison of hypothetical treatment scenarios. *Value Health.* 2014;17(5): 525–536.
24. Regan EA, Hokanson JE, Murphy JR, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD.* 2010;7(1):32–43.
25. Montnemery P, Adelroth E, Heuman K, et al. Prevalence of obstructive lung diseases and respiratory symptoms in southern Sweden. *Respir Med.* 1998;92(12):1337–1345.
26. Kruis AL, Boland MR, Schoonvelde CH, et al. RECODE: design and baseline results of a cluster randomized trial on cost-effectiveness of integrated COPD management in primary care. *BMC Pulm Med.* 2013; 13(1):17.
27. Vestbo J, Anderson W, Coxson HO, et al; ECLIPSE investigators. Evaluation of COPD longitudinally to identify predictive surrogate end-points (ECLIPSE). *Eur Respir J.* 2008;31(4):869–873.
28. 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.
29. Keene ON, Calverley PM, Jones PW, Vestbo J, Anderson JA. Statistical analysis of exacerbation rates in COPD: TRISTAN and ISOLDE revisited. *Eur Respir J.* 2008;32(1):17–24.
30. Al-ani S, Spigt M, Hofset P, Melbye H. Predictors of exacerbations of asthma and COPD during one year in primary care. *Fam Pract.* 2013; 30(6):621–628.
31. Husebo GR, Bakke PS, Aanerud M, et al. Predictors of exacerbations in chronic obstructive pulmonary disease: results from the Bergen COPD cohort study. *PLoS One.* 2014;9(10):e109721.
32. Kilic H, Kokturk N, Sari G, Cakir M. Do females behave differently in COPD exacerbation? *Int J Chron Obstruct Pulmon Dis.* 2015;10: 823–830.
33. Owens GM. Gender differences in health care expenditures, resource utilization, and quality of care. *J Manag Care Pharm.* 2008;14(3):2–6.
34. Kerkhof M, Freeman D, Jones R, Chisholm A, Price DB; Effectiveness Group. Predicting frequent COPD exacerbations using primary care data. *Int J Chron Obstruct Pulmon Dis.* 2015;10:2439–2450.

Supplementary materials

Table S1 Prediction models for total number of exacerbations and number of severe exacerbations using COPDGene data: coefficients (SE) and *p*-values*

	COPDGene ¹			
	Total exacerbations		Severe exacerbations	
	Coefficient (SE)	<i>p</i> -value	Coefficient (SE)	<i>p</i> -value
Intercept	2.06 (1.06)	0.052	-1.68 (1.42)	0.24
Sex (1= female)	0.21 (0.07)	0.001	0.20 (0.09)	0.03
Age (years)	-0.0012 (0.005)	0.79	-0.0037 (0.006)	0.54
FEV ₁ % predicted (in %)	-0.011 (0.002)	<0.001	-0.015 (0.002)	<0.001
Number of exacerbations prior to baseline	-	-	0.77 (0.11) [#]	<0.001
One exacerbation	0.56 (0.08)	<0.001	-	-
Two exacerbations	0.67 (0.12)	<0.001	-	-
Three exacerbations	0.81 (0.17)	<0.001	-	-
Four exacerbations	0.67 (0.23)	0.004	-	-
Five exacerbations	0.52 (0.43)	0.23	-	-
Six or more exacerbations	0.83 (0.24)	<0.001	-	-
BMI <20 kg/m ² (1= yes)	-0.020 (0.13)	0.88	0.26 (0.17)	0.14
History of cardiovascular disease (1= yes)	0.071 (0.11)	0.51	0.23 (0.14)	0.10
SGRQ total score at baseline (in points)	0.015 (0.002)	<0.001	0.023 (0.003)	<0.001
Smoker (1= yes, 0= former)	-0.22 (0.08)	0.006	-0.014 (0.11)	0.90
Cough (1= yes)	0.15 (0.07)	0.043	0.011 (0.10)	0.91
Wheeze (1= yes)	0.32 (0.08)	<0.001	0.29 (0.11)	0.01
Resting O ₂ saturation (in %)	-0.034 (0.01)	0.001	-0.0055 (0.01)	0.69

Notes: [#]Severe exacerbations in the year prior to baseline: yes/no. *Predicted rate can be calculated using the following formula: Predicted rate = e^(intercept + value predictor1 * coefficient predictor1 + value predictor2 * coefficient predictor2 + ...).

Abbreviations: BMI, body mass index; FEV₁, forced expiratory volume in 1 second; SGRQ, St George's Respiratory Questionnaire.

Table S2 Prediction models for total number of exacerbations using OLIN data: coefficients (SE) and *p*-values*

	OLIN ² : Total exacerbations	
	Coefficient (SE)	<i>p</i> -value
Intercept	-0.76 (2.08)	0.71
Sex (1= female)	1.05 (0.43)	0.01
Age (years)	0.0076 (0.03)	0.78
FEV ₁ % predicted (in %)	-0.039 (0.01)	0.002
Number of exacerbations prior to baseline	0.40 (0.10)	<0.001
BMI <20 kg/m ² (1= yes)	0.42 (0.93)	0.65
History of cardiovascular disease (1= yes)	0.43 (0.56)	0.44

Note: *Predicted rate can be calculated using the following formula: Predicted rate = e^(intercept + value predictor1 * coefficient predictor1 + value predictor2 * coefficient predictor2 + ...).

Abbreviations: BMI, body mass index; FEV₁, forced expiratory volume in 1 second; OLIN, Obstructive Lung Disease in Norrbotten.

Table S3 Prediction models for total number of exacerbations and number of severe exacerbations using RECODE data: coefficients (SE) and *p*-values*

	RECODE ³			
	Total exacerbations		Severe exacerbations	
	Coefficient (SE)	<i>p</i> -value	Coefficient (SE)	<i>p</i> -value
Intercept	-0.93 (0.50)	0.06	-4.34 (1.18)	<0.001
Sex (1= female)	0.18 (0.13)	0.17	-0.45 (0.32)	0.16
Age (years)	0.0039 (0.006)	0.52	0.018 (0.01)	0.19
FEV ₁ % predicted (in %)	-0.016 (0.003)	<0.001	-0.018 (0.008)	0.02
Number of exacerbations prior to baseline	0.46 (0.06)	<0.001	1.65 (0.67)	0.01
History of cardiovascular disease (1= yes)	0.057 (0.21)	0.78	-0.44 (0.53)	0.40
SGRQ total score at baseline (in points)	0.018 (0.005)	0.001	0.028 (0.01)	0.008
Treatment group trial (1= yes)	0.072 (0.13)	0.57	-0.038 (0.30)	0.90
Smoker (1= yes, 0= former)	0.069 (0.14)	0.64	0.0 (0.31)	0.11
MRC dyspnea	-0.023 (0.07)	0.73	0.13 (0.14)	0.38
Charlson comorbidity index	-0.020 (0.06)	0.75	0.025 (0.15)	0.87
Physical activity IPAQ (1= low)	-0.21 (0.20)	0.28	-0.018 (0.50)	0.97

Note: *Predicted rate can be calculated using the following formula: Predicted rate = $e^{(\text{intercept} + \text{value predictor1} * \text{coefficient predictor1} + \text{value predictor2} * \text{coefficient predictor2} + \dots)}$.

Abbreviations: FEV₁, forced expiratory volume in 1 second; IPAQ, International Physical Activity Questionnaire; SGRQ, St George's Respiratory Questionnaire.

Table S4 Prediction models for total number of exacerbations and number of severe exacerbations using ECLIPSE data: coefficients (SE) and *p*-values*

	ECLIPSE ⁴			
	Total exacerbations		Severe exacerbations	
	Coefficient (SE)	<i>p</i> -value	Coefficient (SE)	<i>p</i> -value
Intercept	-1.32 (0.37)	<0.001	-3.62 (0.79)	<0.001
Sex (1= female)	0.27 (0.06)	<0.001	0.049 (0.13)	0.71
Age (years)	0.0098 (0.004)	0.02	0.028 (0.009)	0.003
FEV ₁ % predicted (in %)	-0.012 (0.002)	<0.001	-0.024 (0.005)	<0.001
Number of exacerbations prior to baseline	0.32 (0.02)	<0.001	0.69 (0.09)	<0.001
BMI <20 kg/m ² (1= yes)	0.085 (0.09)	0.35	0.36 (0.17)	0.04
History of cardiovascular disease (1= yes)	-0.020 (0.06)	0.74	0.32 (0.12)	0.009
SGRQ total score at baseline (in points)	0.0083 (0.002)	<0.001	0.017 (0.004)	<0.001
Smoker (1= yes, 0= former)	-0.12 (0.006)	0.046	0.11 (0.13)	0.40
MRC dyspnea	0.026 (0.07)	0.69	-0.077 (0.14)	0.58
Other comorbidities (1= yes)	0.002 (0.06)	0.97	-0.14 (0.13)	0.31
6-min walking test (m)	0.0009 (0.0003)	<0.001	-0.0002 (0.0006)	0.70
Fibrinogen (mg/dL)	0.0001 (0.0003)	0.77	0.0009 (0.0006)	0.10

Note: *Predicted rate can be calculated using the following formula: Predicted rate = $e^{(\text{intercept} + \text{value predictor1} * \text{coefficient predictor1} + \text{value predictor2} * \text{coefficient predictor2} + \dots)}$.

Abbreviations: BMI, body mass index; ECLIPSE, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints; FEV₁, forced expiratory volume in 1 second; MRC, Medical Research Council; SGRQ, St George's Respiratory Questionnaire.

Table S5 Prediction models for total number of exacerbations and number of severe exacerbations using UPLIFT data: coefficients (SE) and *p*-values*

	UPLIFT ⁵			
	Total exacerbations		Severe exacerbations	
	Coefficient (SE)	<i>p</i> -value	Coefficient (SE)	<i>p</i> -value
Intercept	-1.17 (0.21)	<0.001	-3.62 (0.39)	<0.001
Sex (1= female)	0.082 (0.05)	0.07	0.019 (0.09)	0.83
Age (years)	0.011 (0.002)	<0.001	0.031 (0.005)	<0.001
FEV ₁ % predicted (in %)	-0.0013 (0.002)	<0.001	-0.032 (0.003)	<0.001
Number of exacerbations prior to baseline	0.23 (0.02)	<0.001	0.052 (0.05)	<0.001
BMI <20 kg/m ² (1= yes)	0.18 (0.06)	0.005	0.61 (0.11)	<0.001
History of cardiovascular disease (1= yes)	-0.015 (0.04)	0.73	0.10 (0.08)	0.20
SGRQ total score at baseline (in points)	0.0074 (0.001)	<0.001	0.014 (0.02)	<0.001
Treatment group trial (1= yes)	-0.18 (0.04)	<0.001	-0.15 (0.07)	0.04
Smoker (1= yes)	0.05 (0.04)	0.26	0.21 (0.08)	0.01
Pack-years	0.002 (0.007)	0.005	0 (0.001)	0.97
Time since diagnosis (years)	0.0047 (0.003)	0.08	0.0068 (0.005)	0.18
Diagnosis of emphysema (1= yes)	-0.032 (0.04)	0.42	0.17 (0.08)	0.03
Charlson comorbidity index	-0.021 (0.03)	0.40	0.028 (0.05)	0.53
ICS at baseline (1= yes)	0.26 (0.04)	<0.001	0.21 (0.08)	0.008

Note: *Predicted rate can be calculated using the following formula: Predicted rate = $e^{(\text{intercept} + \text{value predictor1} * \text{coefficient predictor1} + \text{value predictor2} * \text{coefficient predictor2} + \dots)}$.

Abbreviations: BMI, body mass index; FEV₁, forced expiratory volume in 1 second; ICS, inhaled corticosteroid; SGRQ, St George's Respiratory Questionnaire.

References

1. Regan EA, Hokanson JE, Murphy JR, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD*. 2010;7(1):32–43.
2. Montnemery P, Adelroth E, Heuman K, et al. Prevalence of obstructive lung diseases and respiratory symptoms in southern Sweden. *Respir Med*. 1998;92(12):1337–1345.
3. Kruis AL, Boland MR, Schoonvelde CH, et al. RECODE: design and baseline results of a cluster randomized trial on cost-effectiveness of integrated COPD management in primary care. *BMC Pulm Med*. 2013;13(1):17.
4. Vestbo J, Anderson W, Coxson HO, et al; ECLIPSE investigators. Evaluation of COPD longitudinally to identify predictive surrogate end-points (ECLIPSE). *Eur Respir J*. 2008;31(4):869–873.
5. 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.

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-chronic-obstructive-pulmonary-disease-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.