

# Predicting frequent COPD exacerbations using primary care data

Marjan Kerkhof<sup>1</sup>  
Daryl Freeman<sup>2</sup>  
Rupert Jones<sup>3</sup>  
Alison Chisholm<sup>4</sup>  
David B Price<sup>1,5</sup>

On behalf of the Respiratory Effectiveness Group

<sup>1</sup>Research in Real-Life, Cambridge,  
<sup>2</sup>Mundesley Medical Centre, Norfolk,  
<sup>3</sup>Plymouth University Peninsula School of Medicine and Dentistry, Plymouth,  
<sup>4</sup>Respiratory Effectiveness Group, Cambridge,  
<sup>5</sup>Academic Primary Care, Division of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

**Purpose:** Acute COPD exacerbations account for much of the rising disability and costs associated with COPD, but data on predictive risk factors are limited. The goal of the current study was to develop a robust, clinically based model to predict frequent exacerbation risk.

**Patients and methods:** Patients identified from the Optimum Patient Care Research Database (OPCRD) with a diagnostic code for COPD and a forced expiratory volume in 1 second/forced vital capacity ratio  $<0.7$  were included in this historical follow-up study if they were  $\geq 40$  years old and had data encompassing the year before (predictor year) and year after (outcome year) study index date. The data set contained potential risk factors including demographic, clinical, and comorbid variables. Following univariable analysis, predictors of two or more exacerbations were fed into a stepwise multivariable logistic regression. Sensitivity analyses were conducted for subpopulations of patients without any asthma diagnosis ever and those with questionnaire data on symptoms and smoking pack-years. The full predictive model was validated against 1 year of prospective OPCRD data.

**Results:** The full data set contained 16,565 patients (53% male, median age 70 years), including 9,393 patients without any recorded asthma and 3,713 patients with questionnaire data. The full model retained eleven variables that significantly predicted two or more exacerbations, of which the number of exacerbations in the preceding year had the strongest association; others included height, age, forced expiratory volume in 1 second, and several comorbid conditions. Significant predictors not previously identified included eosinophilia and COPD Assessment Test score. The predictive ability of the full model (*C* statistic 0.751) changed little when applied to the validation data set ( $n=2,713$ ; *C* statistic 0.735). Results of the sensitivity analyses supported the main findings.

**Conclusion:** Patients at risk of exacerbation can be identified from routinely available, computerized primary care data. Further study is needed to validate the model in other patient populations.

**Keywords:** prediction, risk factor, model, validation, FEV<sub>1</sub>

## Introduction

COPD is a serious, debilitating condition that has become a major public health concern and by 2020 is projected to rank fifth in global burden of disease and third in global mortality.<sup>1</sup> For a proportion of patients, COPD is a progressive disease characterized by periodic acute exacerbations of symptoms. These exacerbations pose an immediate threat to patients and also hasten progression of the disease. Exacerbations accelerate decline of lung function, so that patients often fail to return to baseline levels. Although symptoms may last a few days, recovery of lung function can take weeks to months, resulting in prolonged periods of functional limitation and general worsening of quality of life, often with some degree of permanent functional decline. Exacerbations are also associated with substantial risk of hospitalization and death,

Correspondence: David B Price  
Academic Primary Care, Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK  
Tel +44 12 2455 4588  
Fax +44 12 2455 0683  
Email [dprice@rirl.org](mailto:dprice@rirl.org)

as well as considerable economic costs that increase with exacerbation frequency.<sup>1,2</sup>

Preventing exacerbations is therefore a key goal of COPD management<sup>1,3-5</sup> and hence the need to predict who are likely to experience exacerbations. The recent Global Initiative for Chronic Obstructive Lung Disease (GOLD) categories A–D were developed to aid in assessing future risk of exacerbations and performed well in one study;<sup>6,7</sup> however, patient assignment to categories may vary depending on the choice of symptom measure, limiting their applicability.<sup>8</sup> Indices of COPD severity such as the body mass index [BMI], obstruction, dyspnea, and exercise (BODE) index and the age, dyspnea, and obstruction (ADO) index have been used to try to predict future exacerbations among patients with COPD but with only moderate (60%–70%) prediction success; moreover, they require specific data (eg, 6-minute walk test) that may not be routinely available.<sup>9,10</sup> The dyspnea, airflow obstruction, smoking status, and exacerbation frequency (DOSE) index has been shown to predict future exacerbations in a large primary care data set, and the index was stronger than previous exacerbation frequency or the ADO or BODE index.<sup>9,11,12</sup> Other researchers have developed de novo statistical models to identify independent clinical predictors. However, many of these studies included relatively small sample sizes,<sup>10,13–16</sup> patients with severe COPD,<sup>14,17,18</sup> and/or severe outcomes such as hospitalization or death.<sup>13,15,17,19–22</sup>

The goal of the current study was to develop a robust, clinically based predictive model that would encompass all levels of COPD severity as well as moderate or severe exacerbation severity. Such a model could help in earlier targeting of patients for review to optimize drug therapy and other interventions, with the aim of reducing hospital admissions, decline in lung function, and the morbidity and mortality associated with COPD. A secondary objective was to compare the model's predictive value in relation to existing predictive tools.

## Patients and methods

This was a historical follow-up study of patients with COPD identified from the Optimum Patient Care Research Database (OPCRD).<sup>23</sup> The OPCRd is a quality-controlled, longitudinal, respiratory-focused database containing anonymous data from general practices in the UK and has been approved by the Trent Multicentre Research Ethics Committee for clinical research use (approval reference 10/H0405/3), and this study was approved by the Anonymised Data Ethics Protocols and Transparency committee, the independent scientific advisory committee for the OPCRd. Informed consent was not required or possible as we worked with anonymous data, and

this was not an interventional study. However, patients could opt out of having their data used in research. At the time of the study, the OPCRd contained records of >50,000 patients with COPD from >300 UK general practices. The database combines routine data from electronic patient records with linked patient-reported data collected using disease-specific questionnaires. Routine clinical data, including patient demographic characteristics, comorbidities, exacerbation history, modified Medical Research Council (mMRC) score,<sup>24</sup> and current therapy, were extracted from primary care practice management systems. In addition, a proportion of patients with relevant disease codes were invited to complete validated disease assessment questionnaires, sent via a secure mailing house. The questionnaires enabled calculation of the mMRC scores and the COPD Assessment Test (CAT) scores.<sup>25</sup>

The current study was divided into model-building and model-validation components. Patients were eligible for inclusion in the model-building phase if, on or before March 12, 2013, they had at least one recorded eosinophil count, were at least 40 years of age, had been diagnostically coded for COPD, and had a forced expiratory volume in 1 second/forced vital capacity (FEV<sub>1</sub>/FVC) ratio <0.7 recorded within 5 years of their last eosinophil count (defined as the index date). All eligible patients needed to have at least 1 year of observation before (baseline year) and 1 year after (outcome year) the index date. Included patients also needed to have complete data on the candidate predictors analyzed. Those with chronic respiratory diseases other than asthma, such as bronchiectasis, were excluded. The validation cohort consisted of patients with similar eligibility criteria identified between March 2013 and February 2014.

Potentially important variables within the OPCRd were identified from a search of the literature and from expert opinion of the authors:

- *Sociodemographic factors*: sex, age, height, weight, BMI, smoking status
- *Symptom severity*: mMRC dyspnea score, CAT score, number of exacerbations in the previous year
- *Comorbidities*: asthma, eczema, allergic or nonallergic rhinitis, nasal polyps, diabetes mellitus, gastroesophageal reflux disease (GERD), ischemic heart disease, heart failure, anxiety/depression, Charlson comorbidity index
- *Spirometry*: FEV<sub>1</sub> (% predicted), FEV<sub>1</sub>/FVC ratio
- Peripheral blood eosinophilia (defined as  $\geq 500$  cells/ $\mu$ L).

## Model building

To enhance diagnostic specificity and to be consistent with earlier research, as well as with the GOLD cut-point for

risk of future events,<sup>7</sup> we defined the outcome of interest as frequent ( $\geq 2$ ) exacerbations. Exacerbations were defined as either 1) unscheduled hospital admission or accident/emergency attendance for COPD or lower respiratory events, 2) an acute course of oral corticosteroids prescribed with evidence of respiratory review, or 3) antibiotics prescribed with evidence of respiratory review. Where one or more oral corticosteroid course, hospitalization, or antibiotic prescription occurred within a 2-week window, these events were considered to be the result of the same exacerbation.

All analyses were performed using the R statistical package (version 3.0.2). Prior to analysis, continuous variables were evaluated via likelihood ratio test to see if quadratic or cubic transformation improved model fit. Akaike information criteria (AIC) were compared to test whether model fit was improved by categorical or continuous classification.

Univariable logistic regressions were performed to gauge the importance of individual variables and to help define the best form (eg, continuous, categorical) of each variable. However, all potentially important variables were fed into a multiple logistic regression with backward selection of the model having the lowest AIC. The questionnaire variables dealing with symptoms and pack-years were excluded from this analysis owing to small sample size.

By way of sensitivity analysis, the model-building process was repeated for two different subpopulations. Subpopulation 1 consisted of all patients without an overlapping diagnosis of asthma, defining asthma using the sensitive definition of any asthma-related Read code at any time in the data set. The second, smaller subpopulation consisted of those with questionnaire information on symptoms (CAT and mMRC score) and pack-years of smoking.

## Model validation

Calibration plots were performed by comparing observed with predicted risk among 150 groups of  $\sim 110$  patients each. Goodness of fit was judged using both the *C* statistic (area under the curve) and the Hosmer–Lemeshow test.<sup>26</sup> The *C* statistic confidence intervals (CIs) were generated by bootstrapping. External validity was judged by the *C* statistic when the full multivariable model was applied to the validation cohort.

In addition, we assessed the predictive accuracy of the model to predict two or more exacerbations compared with the DOSE index and GOLD categories A–D using the mMRC, together with exacerbations and FEV<sub>1</sub>, to assign categories.<sup>7</sup>

## Results

Approximately 51,000 patients with COPD were identified from the OPCR; 16,565 met all inclusion criteria. The main reasons for exclusion from the study are depicted in Figure 1. The study index dates (ie, dates of last eosinophil count) ranged from 1993 to 2012 (median year, 2009; interquartile range, 2007–2010).

Of the 16,565 patients included in the full population, 9,393 did not have any recorded asthma Read code at any time (subpopulation 1) and 3,713 had questionnaire data for determining CAT score (subpopulation 2). The characteristics of these three populations (Table 1) and the frequencies of COPD exacerbations (Table 2) were similar, with only minor differences. Most patients had moderate COPD, and 92% of lung function measurements were taken within 2 years of the index date (80% within 1 year).

Patient numbers for the categorized mMRC and DOSE index scores and for GOLD category are in Table S1.

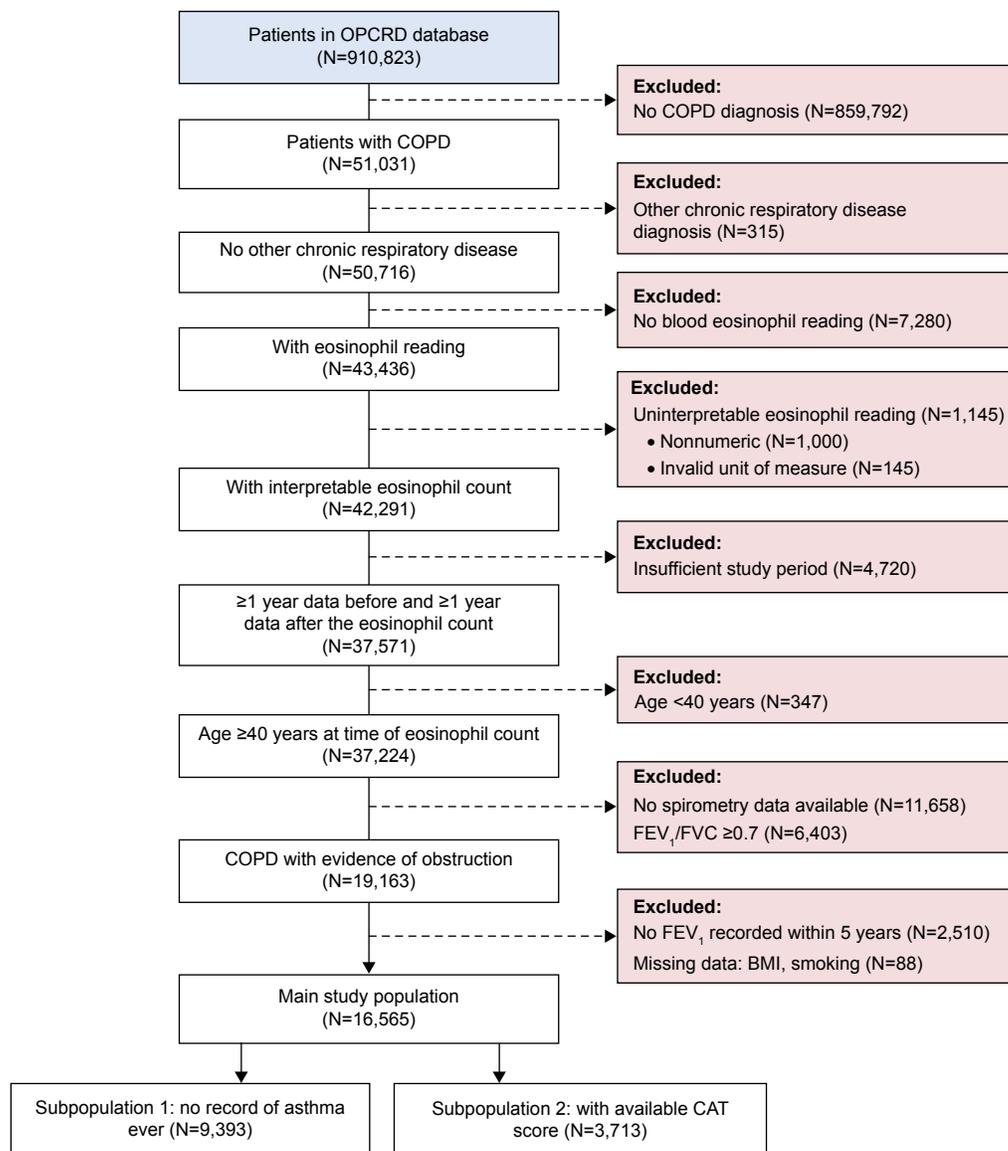
Approximately 20% of the total population had two or more exacerbations in the outcome year (Table 2). Most variables were significant univariable predictors (Table 3).

The final multivariable model contained eleven variables, of which the number of exacerbations in the preceding year had the strongest association. Most other variables were associated with relatively modest odds ratios (ORs) (Table 4). The overall *C* statistic for the model was 0.751 (95% CI 0.742–0.761) (Figure 2). The Hosmer–Lemeshow test had a *P*-value of 0.30, suggesting no significant departures from goodness of fit. The model and a patient example are included in the Supplementary material.

The model developed using the asthma-free subset (subpopulation 1) was very similar to that for the full population (with the obvious exception of asthma) and had similar ORs (Table 5). The smaller subset of patients for whom CAT questionnaire data were available (subpopulation 2) also produced a similar model, although several variables important to the full data set were no longer retained in the subpopulation 2 model. This latter model included two additional variables (CAT score and female sex) not in the full model, but including these variables did not appreciably alter the *C* statistic for the full model. Age was a much stronger risk factor in subpopulation 2 than in the other models (Table 5).

## Model validation

Summary measures of internal validity were unremarkable and suggested adequate fit and predictive ability for the full-population model. Applying the full-population model to the validation data set (N=2,713; baseline characteristics



**Figure 1** Patient selection in the database.

**Abbreviations:** BMI, body mass index; CAT, COPD Assessment Test; OPCR, Optimum Patient Care Research Database; FEV<sub>1</sub>, forced expiratory volume in 1 second; FVC, forced vital capacity.

in Table S2) resulted in a *C* statistic of 0.735 (95% CI 0.713–0.757), suggesting good external validity within the validation cohort (Figure 3).

## Comparison of the model with other indices

The DOSE index score (Table 1) and GOLD group categorization were determined using the mMRC score for 3,558 patients with available data (Table S1). The *C* statistic (95% CI) for a model using the DOSE index was 0.641 (0.617–0.664) and that using the GOLD groups was 0.644 (0.622–0.666) as compared with 0.751 (0.742–0.761) for our full-population model.

## Discussion

Using a large database of routinely collected electronic health records from patients with COPD in the UK, we developed and validated a model incorporating eleven variables that performed well in predicting two or more COPD exacerbations in the subsequent year (*C* statistic of 0.751). Sensitivity analyses in the subpopulations with no asthma ever recorded (*C* statistic 0.742) and with patient-recorded questionnaire data (*C* statistic 0.745) supported the main results. The frequency of exacerbations in the previous year was the major predictor of future exacerbation risk. Our findings provide evidence that routinely collected health care data can be used to predict frequent COPD exacerbations. Moreover, our

**Table 1** Baseline characteristics of all patients and of the two subpopulations

Characteristics	Total population (N=16,565)	Subpopulation 1 (no asthma ever) (N=9,393)	Subpopulation 2 (with symptom data) (N=3,713)
Female sex, n (%)	7,736 (46.7)	4,187 (44.6)	1,623 (43.7)
Age, median (IQR)	70 (63–78)	71 (63–78)	71 (58–78)
Weight (kg), mean (SD)	74.4 (18.3)	74.0 (18.2)	76.3 (18.2)
Height (m), mean (SD)	1.67 (0.10)	1.67 (0.10)	1.68 (0.10)
BMI, mean (SD)	26.7 (5.8)	26.4 (5.7)	27.1 (5.6)
BMI, n (%)			
Underweight (<18.5 kg/m <sup>2</sup> )	859 (5.2)	550 (5.9)	151 (4.1)
Normal (≥18.5 and <25 kg/m <sup>2</sup> )	6,016 (36.3)	3,503 (37.3)	1,293 (34.8)
Overweight (≥25 and <30 kg/m <sup>2</sup> )	5,607 (33.8)	3,151 (33.5)	1,278 (34.4)
Obese (≥30 kg/m <sup>2</sup> )	4,083 (24.6)	2,189 (23.3)	991 (26.7)
Smoking status, n (%)			
Never smoker	1,964 (11.9)	761 (8.1)	455 (12.3)
Ex-smoker	8,875 (53.6)	4,952 (52.7)	2,220 (59.8)
Current smoker	5,726 (34.6)	3,680 (39.2)	1,038 (28.0)
Pack-years of smoking, median (IQR) <sup>a</sup>	n/a	n/a	27.5 (12.5–40.5)
CAT score, median (IQR)	n/a	n/a	17 (11–23)
CAT score ≥10, n (%)	n/a	n/a	3,012 (81.1)
mMRC score available, n (%)			3,558 (95.8)
0–1			1,962 (55.1)
≥2			1,596 (44.9)
DOSE index score available, n (%)			3,558 (95.8)
≤4			3,384 (95.1)
>4			174 (4.9)
FEV <sub>1</sub> /FVC ratio, mean (SD)	54.4 (10.7)	54.6 (10.6)	53.9 (10.6)
FEV <sub>1</sub> % predicted, mean (SD)	57.4 (18.8)	57.7 (18.7)	57.8 (18.1)
GOLD airflow limitation (FEV <sub>1</sub> % predicted), n (%)			
1: mild (≥80%)	1,847 (11.2)	1,053 (11.2)	370 (10.0)
2: moderate (50%–79%)	8,742 (52.8)	5,002 (53.3)	2,069 (55.7)
3: severe (30%–49%)	4,911 (29.6)	2,755 (29.3)	1,071 (28.8)
4: very severe (<30%)	1,065 (6.4)	583 (6.2)	203 (5.5)
Eosinophilia (≥500 cells/μL), n (%)			
Eosinophilia – all patients	1,639 (9.9)	831 (8.8)	336 (9.0)
Eosinophilia – noncurrent smokers	1,128 (6.8)	522 (5.6)	253 (6.8)
Asthma, n (%)	7,172 (43.3)	0	1,587 (42.7)
Eczema, n (%)	3,443 (20.8)	1,748 (18.6)	835 (22.5)
Rhinitis, n (%)			
Allergic	1,245 (7.5)	500 (5.3)	307 (8.3)
Nonallergic	965 (5.8)	507 (5.4)	242 (6.5)
Nasal polyps, n (%)	421 (2.5)	152 (1.6)	94 (2.5)
Diabetes mellitus, n (%)	3,741 (22.6)	2,029 (21.6)	861 (23.2)
GERD, n (%)	1,940 (11.7)	998 (10.6)	446 (12.0)
Ischemic heart disease, n (%)	3,159 (19.1)	1,844 (19.6)	595 (16.0)
Heart failure, n (%)	1,340 (8.1)	806 (8.6)	266 (7.2)
Anxiety or depression, n (%)	5,151 (31.1)	2,822 (30.0)	1,078 (29.0)
Charlson comorbidity index, n (%)			
0	12,802 (77.3)	7,903 (84.1)	2,927 (78.8)
1–4	2,095 (12.6)	577 (6.1)	474 (12.8)
≥5	1,668 (10.1)	913 (9.7)	312 (8.4)

**Notes:** Noncurrent smokers included ex-smokers and never smokers. <sup>a</sup>Pack-years data were available for 3,107 patients.

**Abbreviations:** BMI, body mass index; CAT, COPD Assessment Test; DOSE, dyspnea, airway obstruction, smoking status, exacerbations; FEV<sub>1</sub>/FVC, forced expiratory volume in 1 second/forced vital capacity; GERD, gastroesophageal reflux disease; GOLD, Global Initiative for Chronic Obstructive Lung Disease; IQR, interquartile range; mMRC, modified Medical Research Council; n/a, not applicable; SD, standard deviation.

**Table 2** Number of exacerbations in baseline and outcome years

Number of exacerbations	Total population (N=16,565)	Subpopulation 1 (no asthma ever) (N=9,393)	Subpopulation 2 (with symptom data) (N=3,713)
Baseline year, n (%)			
0	8,783 (53.0)	5,406 (57.6)	2,003 (53.9)
1	4,101 (24.8)	2,277 (24.2)	912 (24.6)
2	1,940 (11.7)	950 (10.1)	420 (11.3)
3	900 (5.4)	422 (4.5)	206 (5.5)
≥4	841 (5.1)	338 (3.6)	172 (4.6)
Outcome year, n (%)			
0	9,347 (56.4)	5,640 (60.0)	2,096 (56.5)
1	3,973 (24.0)	2,245 (23.9)	906 (24.4)
2	1,722 (10.4)	844 (9.0)	398 (10.7)
3	754 (4.6)	339 (3.6)	151 (4.1)
≥4	769 (4.6)	325 (3.5)	162 (4.4)

**Table 3** Univariable predictors of two or more COPD exacerbations in the outcome year in the total population data set (N=16,565)

Variable	Odds ratio (95% CI)	P-value
Female sex	1.25 (1.16–1.35)	9×10 <sup>-9</sup>
Age (per 10 years)	1.71 (1.14–2.56)	0.009
Age <sup>2</sup> (per 10 years)	0.96 (0.93–0.98)	0.003
Height (per 10 cm)	0.88 (0.85–0.92)	6×10 <sup>-10</sup>
Body mass index		
Normal	1.00	0.30
Underweight	1.10 (0.92–1.31)	0.60
Overweight	1.02 (0.93–1.12)	0.45
Obese	1.04 (0.94–1.15)	
Smoking status		
Never	1.00	0.23
Former	1.08 (0.95–1.22)	0.48
Current	0.95 (0.84–1.09)	
Exacerbations in the baseline year		
0	1.00	3×10
1	2.55 (2.30–2.84)	5×10
2	4.86 (4.31–5.47)	7×10
3	8.34 (7.17–9.68)	8×10
≥4	21.05 (17.90–24.75)	
Asthma	1.67 (1.55–1.81)	7×10 <sup>-39</sup>
Rhinitis		
Allergic	1.12 (0.97–1.29)	0.13
Nonallergic	1.60 (1.38–1.85)	5×10 <sup>-10</sup>
Nasal polyps	1.49 (1.19–1.85)	0.0004
GERD	1.37 (1.22–1.53)	3×10 <sup>-8</sup>
Anxiety/depression	1.32 (1.22–1.44)	8×10 <sup>-12</sup>
Eczema	1.17 (1.06–1.28)	0.001
Diabetes	1.12 (1.02–1.22)	0.01
Ischemic heart disease	1.12 (1.02–1.23)	0.02
Heart failure	1.12 (0.98–1.29)	0.09
Charlson comorbidity index		
0	1.00	
1	1.41 (1.27–1.58)	5×10 <sup>-10</sup>
≥2	1.22 (1.08–1.38)	0.002
FEV <sub>1</sub> % (per 10% decrease)	1.15 (1.12–1.17)	5×10 <sup>-36</sup>
FEV <sub>1</sub> /FVC (per 10% decrease)	1.17 (1.13–1.22)	4×10 <sup>-19</sup>
Blood eosinophilia		
Noncurrent smokers	1.43 (1.25–1.64)	7×10 <sup>-7</sup>
Current smokers	0.89 (0.70–1.12)	0.31

**Note:** Noncurrent smokers included ex-smokers and never smokers.

**Abbreviations:** CI, confidence interval; FEV<sub>1</sub>/FVC, forced expiratory volume in 1 second/forced vital capacity; GERD, gastroesophageal reflux disease.

model performed better for predicting COPD exacerbations when applied to our heterogeneous study population than models using the DOSE index or GOLD groups calculated using the mMRC score.

Many other predictive studies have focused on risk of death or hospitalization, often among patients with severe COPD.<sup>13,15,17,19–21</sup> A strength of the current study is the inclusion of all individuals with COPD in a general population and all subsequent exacerbations, regardless of whether the exacerbation required hospitalization.

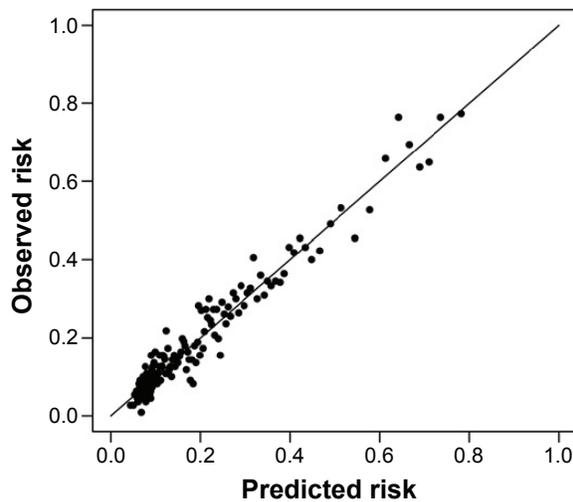
Exacerbation rates were relatively low in the study, with >50% of patients having no acute exacerbation in either the baseline or the outcome years, possibly a result of the broadly inclusive eligibility criteria that produced a

**Table 4** Significant multivariable predictors of two or more COPD exacerbations in the outcome year in the total population data set (N=16,565)

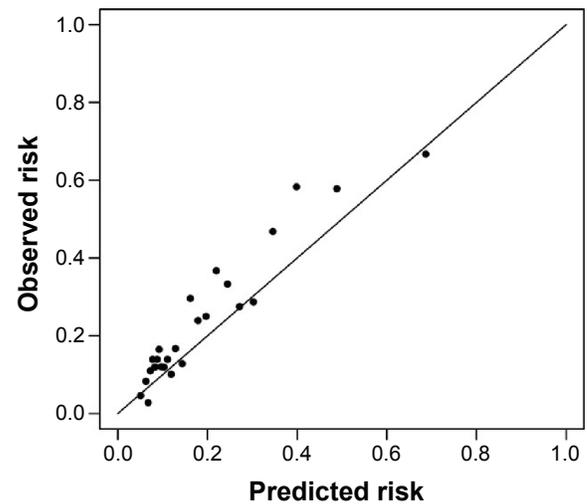
Covariate	Odds ratio (95% CI)
Exacerbations in the baseline year	
0	1.00
1	2.42 (2.18–2.69)
2	4.39 (3.89–4.95)
3	7.28 (6.25–8.48)
≥4	17.83 (15.12–21.03)
FEV <sub>1</sub> % predicted (per 10% decrease)	1.10 (1.07–1.12)
Age (per 10 years)	1.43 (0.92–2.23)
Age <sup>2</sup> (per 10 years)	0.97 (0.93–1.00)
Height (per 10 cm)	0.89 (0.85–0.93)
Eosinophilia in noncurrent smokers	1.29 (1.10–1.51)
Asthma	1.34 (1.23–1.46)
Nonallergic rhinitis	1.35 (1.15–1.59)
Nasal polyps	1.39 (1.09–1.78)
Ischemic heart disease	1.12 (1.01–1.25)
Anxiety or depression	1.11 (1.02–1.22)
GERD	1.18 (1.05–1.34)
Model C statistic (95% CI)	0.751 (0.742–0.761)

**Note:** Noncurrent smokers included ex-smokers and never smokers.

**Abbreviations:** CI, confidence interval; FEV<sub>1</sub>, forced expiratory volume in 1 second; GERD, gastroesophageal reflux disease.



**Figure 2** Calibration plot of observed versus predicted risk using the full developmental model (N=16,565).



**Figure 3** Calibration plot (25 groups of 108–109 observations) of the observed versus predicted risk after applying the model to the validation cohort (N=2,713).

distribution of COPD severities, from mild to very severe, within the study population;<sup>27</sup> primary care COPD populations are recognized as having lower rates of exacerbations than patients enrolled in clinical trials.<sup>28</sup> Moreover, the relatively low rate of exacerbations seen in this study, as compared with past research suggesting mean annual rates of 0.8 in mild COPD and 1.2–2.0 in moderate to very

severe COPD,<sup>27</sup> may be a reflection of changes in recent years, including better identification of milder COPD, with spirometry being broadly undertaken, and more focused COPD management in UK primary care.<sup>29</sup>

A value close to 1 for the *C* statistic indicates that a model has excellent discriminatory power.<sup>30</sup> While a *C* statistic of 0.75 for our model indicates modest predictive ability, the

**Table 5** Significant multivariable predictors of two or more COPD exacerbations in the outcome year among subpopulations

Covariate	OR (95% CI)	
	Subpopulation 1 (no asthma ever) (N=9,393)	Subpopulation 2 (with symptom data) (N=3,713)
Exacerbations in the baseline year		
0	1.00	1.00
1	2.34 (2.02–2.71)	2.17 (1.73–2.71)
2	4.46 (3.75–5.30)	4.26 (3.30–5.51)
3	7.86 (6.31–9.79)	6.49 (4.71–8.93)
≥4	18.18 (14.21–23.26)	15.11 (10.57–21.59)
FEV <sub>1</sub> , % predicted (per 10% decrease)	1.09 (1.06–1.14)	1.05 (1.00–1.11)
Age (per 10 years)	1.59 (0.81–3.11)	5.45 (1.77–16.78)
Age <sup>2</sup> (per 10 years)	0.96 (0.91–1.01)	0.89 (0.82–0.96)
Height (per 10 cm)	0.86 (0.80–0.91)	NI
Eosinophilia in noncurrent smokers	1.41 (1.11–1.79)	1.43 (1.04–1.98)
Asthma	NI	1.19 (1.00–1.43)
Nonallergic rhinitis	1.45 (1.14–1.83)	NI
Nasal polyps	1.52 (1.00–2.32)	1.95 (1.18–3.20)
Heart failure	1.30 (1.06–1.60)	NI
Diabetes	1.11 (0.97–1.28)	NI
Anxiety or depression	1.17 (1.03–1.33)	NI
GERD	1.22 (1.02–1.46)	NI
Female sex	NI	1.31 (1.09–1.57)
CAT score (per 10 units)	NI	1.28 (1.15–1.42)
Model <i>C</i> statistic	0.742 (0.728–0.756)	0.745 (0.724–0.766)

**Note:** Noncurrent smokers included ex-smokers and never smokers.

**Abbreviations:** CAT, COPD Assessment Test; CI, confidence interval; FEV<sub>1</sub>, forced expiratory volume in 1 second; GERD, gastroesophageal reflux disease; NI, not included in the model; OR, odds ratio.

results of the current study compare favorably with those of the earlier studies focused on predicting exacerbations. Miravittles et al<sup>31</sup> performed a cross-sectional assessment of frequent ( $\geq 1$  per year) exacerbation occurrence among 627 ambulatory patients with COPD. Significant covariates included age, FEV<sub>1</sub>, and chronic mucus hypersecretion, but none of these were particularly strong risk factors (OR for hypersecretion 1.54), and the predictive ability of the model was marginal (*C* statistic 0.6).

A substantial number of prospective studies have also attempted to predict COPD exacerbations. Niewoehner et al<sup>18</sup> followed 1,829 veterans for 6 months to assess the risk of either COPD exacerbation or COPD hospitalization. Significant independent predictors for exacerbation included older age, FEV<sub>1</sub>, productive cough, previous hospitalization, and medications used previously. However, all patients had moderate to severe COPD, and the short follow-up may have limited the predictive ability of the model, which was itself inferior (*C* statistic 0.67). Hurst et al<sup>32</sup> followed 2,138 patients with COPD for 3 years to assess risk of COPD exacerbations requiring antibiotics, corticosteroids, or hospitalization. Significant multivariable predictors of two or more exacerbations included previous exacerbation, FEV<sub>1</sub>, history of GERD, increased white cell count, and respiratory health status. Predictive ability of the full model was not reported. Bertens et al<sup>16</sup> followed 243 patients with COPD for 24 months to assess the risk of exacerbation occurrence. Their multivariable model identified FEV<sub>1</sub>, smoking pack-years, history of vascular disease, and previous exacerbations (as a dichotomous variable, yes/no) as significant predictors. The model *C* statistic (0.75) suggested good predictive ability, but the small sample raises issues as to general applicability of these findings, which are mirrored by a marginal external validity (validation *C* statistic 0.66). Bowler et al<sup>33</sup> followed 3,804 patients with COPD for an average of 3 years, identifying ten significant exacerbation predictors, especially FEV<sub>1</sub>, St George severity score, and exacerbation in the previous year. However, none were strong risk factors (OR 1.19 per exacerbation in the previous year), and information on overall predictive ability and external validation was not provided.

The largest prospective study conducted to date included ~59,000 patients with COPD from a primary care database and followed them for 1 year. Multivariable logistic regression identified many significant predictors of two or more exacerbations, including previous exacerbation, air-flow, level of dyspnea, female sex, and various comorbidities (eg, heart failure, renal disease, anxiety, and asthma). Unfortunately, predictive ability of the model was not reported.<sup>34</sup>

Most recently, Make et al<sup>35</sup> followed 3,141 patients from several drug trials with a history of  $\geq 1$  COPD exacerbation in the previous year in order to predict the 6-month risk of an exacerbation requiring corticosteroids or emergency/hospital visit. Independent predictors from a multivariable model included number of maintenance medications, inhaler use, exacerbations in the previous year, FEV<sub>1</sub>/FVC ratio, female sex, and respiratory health status. The *C* statistic of 0.67 suggested only moderate predictive ability.

Our study identified the number of exacerbations in the previous year as a significant predictive factor, which was borne out by most of the studies cited earlier. Bowler et al<sup>33</sup> and Make et al<sup>35</sup> reported only moderately increased risk for previous exacerbation, although the latter study used one previous exacerbation as the reference value (rather than none) in patients with at least one previous exacerbation. However, Bertens et al<sup>16</sup> reported an OR of 5.07 (95% CI 2.55–10.07) for at least one exacerbation in the previous year, and Hurst et al<sup>32</sup> reported an OR of 5.72 (95% CI 4.47–7.31) for this same measure. Müllerová et al<sup>34</sup> reported an exposure–response relationship, with one previous exacerbation associated with an OR of 3.31 (95% CI 3.12–3.51) and two or more associated with an OR of 13.64 (12.67–14.68). The findings from these latter three studies are consistent with those reported in the current study. Furthermore, our results and those reported by Müllerová et al<sup>34</sup> suggest that risk increases with increasing number of previous exacerbations, highlighting the importance of obtaining detailed information on this variable.

Neither our multivariable model nor most other predictive models have identified an independent association between smoking and frequent exacerbations in a broad population of patients with COPD. This suggests that the impact of this variable is dependent on inclusion criteria frequently applied to select patients with COPD for controlled trials. Any association with smoking may also be ameliorated by a so-called “healthy smoker effect”, in which those with poorer lung function or frequent exacerbations tend to quit smoking, whereas less severely affected patients do not.<sup>18</sup> This suggests that there is a phenotypical propensity to frequent exacerbation that is somewhat independent of other risk factors.<sup>32,34</sup>

A unique finding of the current study was the significance of blood eosinophilia as an independent predictor. This is consistent with research showing that eosinophils are present in 20%–40% of sputum samples from patients with stable COPD and that airway eosinophilia increases during exacerbation episodes.<sup>36</sup> This variable had only a moderate

OR of ~1.3, and only among patients not currently smoking, so there may be somewhat limited applicability for predicting exacerbations in general. However, the relatively weak association with eosinophilia may reflect more active COPD management among this population compared with others. We conducted a post hoc sensitivity analysis to evaluate whether the association between blood eosinophilia and the risk of two or more exacerbations would be relevantly different after excluding the 17% ( $n=2,785$ ) of patients with blood eosinophil counts measured at an exacerbation. The results were not relevantly different after excluding these measurements from the analyses (OR 1.26 [95% CI 1.04–1.52] vs 1.29 [1.10–1.51] for the full population).

We also identified CAT score as a significant predictor for subpopulation 2, which to the best of our knowledge has not been reported elsewhere. Indices of COPD severity such as the BODE index have been significantly linked to future exacerbations,<sup>9,37</sup> but these are more complex than the simple CAT survey.

A major strength of the current study is the sample size of >16,000 patients, which is much larger than most other predictive studies. In fact, the population in the current study was more than fivefold larger than all but that of Müllerová et al.<sup>34</sup> However, it is likely that some of the variables retained in our final model reached significance primarily because of this large sample rather than because of strong biological importance. Müllerová et al<sup>34</sup> similarly reported relatively weak OR for many of their predictors but without information on the overall model predictability or the predictive power of individual variables.

We identified a number of comorbid predictors, including heart disease, GERD, and other respiratory conditions. Conditions such as GERD and heart disease have been identified in other follow-up studies,<sup>32–34</sup> whereas risk factors such as nasal polyps and rhinitis appear unique to the current study. These findings highlight the potentially complex relationship that may exist between COPD and other conditions. Although these comorbidities were not strong risk factors (OR 1.1–1.4), their importance would likely improve in a model of newly diagnosed COPD patients without history of exacerbation. Of course, some of these significant predictors may have been driven by the large sample size.

Asthma is recognized as an important comorbid condition that increases disability and risk of exacerbation among those with COPD.<sup>38,39</sup> Asthma was also a significant independent predictor in the current study. However, the OR for asthma was of similar magnitude as that for several other variables, and excluding patients with any record of asthma did not

appreciably change the model. Such findings may reflect an overuse of the asthma “label”, especially in the past. We applied a very sensitive definition to select the subpopulation of patients without overlapping asthma, namely, the recording of any Read code that could indicate the general practitioner (GP) was considering asthma, which explains the relatively high proportion of patients reported with COPD and concomitant asthma. We cannot exclude the possibility of some patients being wrongly diagnosed with asthma by the GP or, conversely, of some patients having undiagnosed asthma.

Differences between the current study and previous clinical research may partially reflect the populations studied. Clinical trials typically enroll restricted populations with more severe disease, often with a greater frequency of exacerbation at baseline.<sup>16,35</sup> The current population better reflects the broader landscape of patients with COPD treated in routine primary care practice in the UK. Nonetheless, we cannot rule out the possibility of selection bias. For example, we required an FEV<sub>1</sub>/FVC ratio of <0.7 for study inclusion. Because spirometry is not universally available in primary care settings, 11,658 of 37,224 (31%) otherwise potentially eligible patients did not have spirometry results and hence were excluded. With regard to the requirement for blood eosinophil count, full blood count measurements are very common among patients with COPD and were available for 86% of patients evaluated (43,436 of 50,716 patients with COPD and no other chronic respiratory disease).

Our aim was to evaluate the predictive value of routinely collected data. Although other parameters of eosinophilic inflammation, such as sputum eosinophils or exhaled nitric oxide, may have improved the predictive performance of the model, these measurements are generally unavailable in general practice and hence were not included in our model. The current study used electronic records from primary care providers, which are a readily available data source. However, it is possible that outcomes such as hospital and emergency admissions may be underrepresented in the data.

The current study was validated both internally and externally, with good concordance when the multivariable model was applied to an external sample. However, this does not guarantee universal generalizability given that our external sample arose from the same patient population as the sample used for model development. Further study is needed to validate the model in other patient populations as well.

## Conclusion

Routine, electronic medical record data available from most GP clinical systems can be used to identify patients with

COPD at risk of two or more exacerbations the subsequent year. Our model could be used to profile patients with COPD, or to underpin decision support tools, in general practice. The number of exacerbations in the preceding year showed a strong exposure–response relationship, highlighting the importance of detailed information on patients' exacerbation history. The findings also suggest that CAT score and eosinophilia may be convenient markers of future exacerbation, at least in some populations.

## Acknowledgments

This study was funded by an unrestricted grant from the Respiratory Effectiveness Group (REG; [www.effectivenessevaluation.org](http://www.effectivenessevaluation.org)). Access to data from the Optimum Patient Care Research Database was co-funded by Research in Real-Life Ltd (RiRL, Cambridge, UK). The authors would like to thank Dr John Bukowski of WordsWorld Consulting for editorial assistance drafting this manuscript. Additional editorial support was provided by Elizabeth V Hillyer, DVM, funded by RiRL.

## Author contributions

Drs Freeman and Price were responsible for study conception and design. Dr Kerkhof was also involved with study design and data analysis. All authors were involved with interpretation of the data, writing, and revision of the paper and decision to submit the paper for publication.

## Disclosure

MK is an employee of Research in Real-Life, which conducted this study and which has conducted paid research in respiratory disease on behalf of the following other organizations in the past 5 years: Aerocrine, AKL Ltd, Almirall, AstraZeneca, Boehringer Ingelheim, Chiesi, GlaxoSmithKline, Meda, Mundipharma, Napp, Novartis, Orion, Takeda, Teva, and Zentiva, a Sanofi company.

DF is a director of Respiratory Education Alliance, a company providing educational packages in the Middle and Far East, and has received fees for lectures/professional consultation from AZ, BI, GSK, Mundipharma, Novartis, and TEVA and has received educational support from Napp, Almirall, BI, and Novartis.

RJ reports personal fees from Almirall, AstraZeneca, Boehringer Ingelheim, Chiesi, Cipla GSK; grants, personal fees, nonfinancial support from Novartis; personal fees and nonfinancial support from Napp/Mundipharma.

AC has no relevant conflicts to declare.

DBP has Board Membership with Aerocrine, Almirall, Amgen, AstraZeneca, Boehringer Ingelheim, Chiesi, Meda,

Mundipharma, Napp, Novartis, and Teva. Consultancy: Almirall, Amgen, AstraZeneca, Boehringer Ingelheim, Chiesi, GlaxoSmithKline, Meda, Mundipharma, Napp, Novartis, Pfizer, Teva, and Zentiva; grants/grants pending with UK National Health Service, British Lung Foundation, Aerocrine, AstraZeneca, Boehringer Ingelheim, Chiesi, Eli Lilly, GlaxoSmithKline, Meda, Merck, Mundipharma, Novartis, Orion, Pfizer, Respiratory Effectiveness Group, Takeda, Teva, and Zentiva; payments for lectures/speaking: Almirall, AstraZeneca, Boehringer Ingelheim, Chiesi, Cipla, GlaxoSmithKline, Kyorin, Meda, Merck, Mundipharma, Novartis, Pfizer, SkyePharma, Takeda, and Teva; payment for manuscript preparation: Mundipharma and Teva; patents (planned, pending, or issued): AKL Ltd; payment for the development of educational materials: GlaxoSmithKline, Novartis; stock/stock options: shares in AKL Ltd that produces phytopharmaceuticals and owns 80% of Research in Real-Life Ltd and its subsidiary social enterprise Optimum Patient Care; received payment for travel/accommodations/meeting expenses from Aerocrine, Boehringer Ingelheim, Mundipharma, Napp, Novartis, and Teva; funding for patient enrollment or completion of research: Almirall, Chiesi, Teva, and Zentiva; peer reviewer for grant committees: Medical Research Council (2014), Efficacy and Mechanism Evaluation programme (2012), HTA (2014); and received unrestricted funding for investigator-initiated studies from Aerocrine, AKL Ltd, Almirall, Boehringer Ingelheim, Chiesi, Meda, Mundipharma, Napp, Novartis, Orion, Takeda, Teva, and Zentiva. The authors report no other conflicts of interest in this work.

## References

1. Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med.* 2013;187(4):347–365.
2. Punekar YS, Shukla A, Mullerova H. COPD management costs according to the frequency of COPD exacerbations in UK primary care. *Int J Chron Obstruct Pulmon Dis.* 2014;9:65–73.
3. Croxton TL, Weinmann GG, Senior RM, Wise RA, Crapo JD, Buist AS. Clinical research in chronic obstructive pulmonary disease: needs and opportunities. *Am J Respir Crit Care Med.* 2003;167(8):1142–1149.
4. Celli B, Cross S, Grossman R, et al. Improving the care of COPD patients – suggested action points by the COPD exacerbations taskforce for reducing the burden of exacerbations of COPD. *Prim Care Respir J.* 2006;15(3):139–142.
5. Decramer M, Janssens W, Miravittles M. Chronic obstructive pulmonary disease. *Lancet.* 2012;379(9823):1341–1351.
6. Lange P, Marott JL, Vestbo J, et al. Prediction of the clinical course of chronic obstructive pulmonary disease, using the new GOLD classification: a study of the general population. *Am J Respir Crit Care Med.* 2012;186(10):975–981.
7. Global Initiative for Chronic Obstructive Lung Disease (GOLD). *Global Strategy for the Diagnosis, Management and Prevention of COPD, 2015*; 2015. Available from: <http://www.goldcopd.org/>. Accessed July 7, 2015.

8. Han MK, Muellerova H, Curran-Everett D, et al. GOLD 2011 disease severity classification in COPDGene: a prospective cohort study. *Lancet Respir Med*. 2013;1(1):43–50.
9. Motegi T, Jones RC, Ishii T, et al. A comparison of three multidimensional indices of COPD severity as predictors of future exacerbations. *Int J Chron Obstruct Pulmon Dis*. 2013;8:259–271.
10. Faganello MM, Tanni SE, Sanchez FF, Pelegrino NR, Lucheta PA, Godoy I. BODE index and GOLD staging as predictors of 1-year exacerbation risk in chronic obstructive pulmonary disease. *Am J Med Sci*. 2010;339(1):10–14.
11. 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.
12. Jones R, Price D, Chavannes N, et al. A comparison of multi-component indices of COPD severity in primary care: an UNLOCK study from the IPCRG. *Eur Respir J*. 2013;42(suppl 57):2006.
13. Garcia-Aymerich J, Monsó E, Marrades RM, et al; EFRAM Investigators. Risk factors for hospitalization for a chronic obstructive pulmonary disease exacerbation. EFRAM study. *Am J Respir Crit Care Med*. 2001;164(6):1002–1007.
14. Foreman MG, DeMeo DL, Hersh CP, Reilly JJ, Silverman EK. Clinical determinants of exacerbations in severe, early-onset COPD. *Eur Respir J*. 2007;30(6):1124–1130.
15. Bahadori K, FitzGerald JM, Levy RD, Fera T, Swiston J. Risk factors and outcomes associated with chronic obstructive pulmonary disease exacerbations requiring hospitalization. *Can Respir J*. 2009;16(4):e43–e49.
16. 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.
17. Fan VS, Ramsey SD, Make BJ, Martinez FJ. Physiologic variables and functional status independently predict COPD hospitalizations and emergency department visits in patients with severe COPD. *COPD*. 2007;4(1):29–39.
18. Niewoehner DE, Lokhnygina Y, Rice K, et al. Risk indexes for exacerbations and hospitalizations due to COPD. *Chest*. 2007;131(1):20–28.
19. Austin PC, Stanbrook MB, Anderson GM, Newman A, Gershon AS. Comparative ability of comorbidity classification methods for administrative data to predict outcomes in patients with chronic obstructive pulmonary disease. *Ann Epidemiol*. 2012;22(12):881–887.
20. Rowe BH, Villa-Roel C, Guttman A, et al. Predictors of hospital admission for chronic obstructive pulmonary disease exacerbations in Canadian emergency departments. *Acad Emerg Med*. 2009;16(4):316–324.
21. Stiell IG, Clement CM, Aaron SD, et al. Clinical characteristics associated with adverse events in patients with exacerbation of chronic obstructive pulmonary disease: a prospective cohort study. *CMAJ*. 2014;186(6):E193–E204.
22. Liu D, Peng SH, Zhang J, Bai SH, Liu HX, Qu JM. Prediction of short term re-exacerbation in patients with acute exacerbation of chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis*. 2015;10:1265–1273.
23. Optimum Patient Care Research Database (OPCRD). Available from: <http://optimumpatientcare.org/>. Accessed July 7, 2015.
24. Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax*. 1999;54(7):581–586.
25. Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidyn N. Development and first validation of the COPD Assessment Test. *Eur Respir J*. 2009;34(3):648–654.
26. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York: Wiley; 1989.
27. Hoogendoorn M, Feenstra TL, Hoogenveen RT, Al M, Molken MR. Association between lung function and exacerbation frequency in patients with COPD. *Int J Chron Obstruct Pulmon Dis*. 2010;5:435–444.
28. Kruis AL, Ställberg B, Jones RC, et al. Primary care COPD patients compared with large pharmaceutically-sponsored COPD studies: an UNLOCK validation study. *PLoS One*. 2014;9(3):e90145.
29. Jones RC, Price D, Ryan D, et al; Respiratory Effectiveness Group. Opportunities to diagnose chronic obstructive pulmonary disease in routine care in the UK: a retrospective study of a clinical cohort. *Lancet Respir Med*. 2014;2(4):267–276.
30. Pencina MJ, D'Agostino RB Sr. Evaluating discrimination of risk prediction models: the C statistic. *JAMA*. 2015;314(10):1063–1064.
31. Miravittles M, Guerrero T, Mayordomo C, Sanchez-Agudo L, Nicolau F, Segu JL. Factors associated with increased risk of exacerbation and hospital admission in a cohort of ambulatory COPD patients: a multiple logistic regression analysis. The EOLO Study Group. *Respiration*. 2000;67(5):495–501.
32. 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.
33. Bowler RP, Kim V, Regan E, et al; COPDGene Investigators. Prediction of acute respiratory disease in current and former smokers with and without COPD. *Chest*. 2014;146(4):941–950.
34. Müllerová 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.
35. 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.
36. Saha S, Brightling CE. Eosinophilic airway inflammation in COPD. *Int J Chron Obstruct Pulmon Dis*. 2006;1(1):39–47.
37. Puhlan 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.
38. Hardin M, Silverman EK, Barr RG, et al; COPDGene Investigators. The clinical features of the overlap between COPD and asthma. *Respir Res*. 2011;12:127.
39. Papaiwannou A, Zarogoulidis P, Porpodis K, et al. Asthma-chronic obstructive pulmonary disease overlap syndrome (ACOS): current literature review. *J Thorac Dis*. 2014;6(suppl 1):S146–S151.

## Supplementary materials

The formula: Risk of  $\geq 2$  COPD exacerbations within the next 12 months =  $1/(1 + \exp(-0.7306 + 0.8840 \times 1 \text{ previous exacerbation in last 12 months} + 1.4786 \times 2 \text{ previous exacerbations in last 12 months} + 1.9857 \times 3 \text{ previous exacerbations in last 12 months} + 2.8811 \times \geq 4 \text{ previous exacerbations in last 12 months} - 0.0093 \times \text{FEV}_1 \text{ predicted} + 0.0360 \times \text{age} - 0.0004 \times \text{age}^2 - 1.2194 \times \text{height (in meter)} + 0.2518 \times (\text{blood eosinophil count} \geq 400/\mu\text{L in a patient who is not currently smoking}) + 0.2953 \times \text{any evidence of asthma} \times 0.3018 \times \text{history of nonallergic rhinitis} + 0.3298 \times \text{history of nasal polyps} + 0.1164 \times \text{history of ischemic heart disease} + 0.1071 \times \text{history of anxiety of depression} + 0.1689 \times \text{history of GERD})$ .

Example: a person aged 70 years currently smoking with height of 1.80 m and FEV<sub>1</sub> of 60% of predicted, without a history of previous exacerbations in the last 12 months and no history of comorbidities, has a calculated risk of 0.064 (6.4%) of two or more exacerbations in the next year.

**Table S1** Baseline mMRC scores, DOSE index scores, and GOLD groups based on mMRC score

	Subpopulation 2 (with symptom data) (N=3,558)
mMRC score	
0	558 (15.7)
1	1,404 (39.5)
2	802 (22.5)
3	539 (15.1)
4	255 (7.2)
DOSE index score	
0	846 (23.8)
1	1,070 (30.1)
2	711 (20.0)
3	469 (13.2)
4	288 (8.1)
>4	174 (4.9)
GOLD category	
A	1,213 (34.1)
B	661 (18.6)
C	749 (21.1)
D	935 (26.3)

**Notes:** mMRC data (hence DOSE index scores and GOLD category based on mMRC)<sup>1-3</sup> were available for 3,558 (95.8%) of 3,713 patients in subpopulation 2. Data are presented as n (% of 3,558). GOLD categories were calculated using mMRC, exacerbations, and FEV<sub>1</sub>.

**Abbreviations:** DOSE, dyspnea, airway obstruction, smoking status, exacerbations; GOLD, Global Initiative for Chronic Obstructive Lung Disease; mMRC, modified Medical Research Council; FEV<sub>1</sub>, forced expiratory volume in 1 second.

**Table S2** Baseline characteristics of the validation cohort

Characteristics	Total population (N=2,713)
Female sex, n (%)	1,237 (45.6)
Age, median (IQR)	71 (64–79)
Weight (kg), mean (SD)	74.0 (18.2)
Height (m), mean (SD)	1.67 (0.10)
Body mass index, mean (SD)	26.6 (5.7)
Body mass index, n (%)	
Underweight (<18.5 kg/m <sup>2</sup> )	157 (5.8)
Normal ( $\geq 18.5$ and <25 kg/m <sup>2</sup> )	1,006 (37.1)
Overweight ( $\geq 25$ and <30 kg/m <sup>2</sup> )	904 (33.3)
Obese ( $\geq 30$ kg/m <sup>2</sup> )	646 (23.8)
Smoking status, n (%)	
Never smoker	341 (12.6)
Ex-smoker	1,365 (50.3)
Current smoker	1,007 (37.1)
FEV <sub>1</sub> /FVC ratio, mean (SD)	54.5 (11.0)
FEV <sub>1</sub> % predicted, mean (SD)	57.4 (19.9)
GOLD airflow limitation (FEV <sub>1</sub> % predicted), n (%)	
1: mild ( $\geq 80\%$ )	331 (12.2)
2: moderate (50%–79%)	1,345 (49.6)
3: severe (30%–49%)	830 (30.6)
4: very severe (<30%)	207 (7.6)
Eosinophilia ( $\geq 500$ cells/ $\mu\text{L}$ ), n (%)	
Eosinophilia – all patients	252 (9.3)
Eosinophilia – noncurrent smokers	199 (7.3)
Asthma, n (%)	1,335 (49.2)
Eczema, n (%)	560 (20.6)
Rhinitis, n (%)	
Allergic	225 (8.3)
Nonallergic	171 (6.3)
Nasal polyps, n (%)	59 (2.2)
Ischemic heart disease, n (%)	655 (24.1)
Anxiety or depression, n (%)	908 (33.5)
GERD	358 (13.2)

**Abbreviations:** IQR, interquartile range; SD, standard deviation; FEV<sub>1</sub>/FVC, forced expiratory volume in 1 second/forced vital capacity; GOLD, Global Initiative for Chronic Obstructive Lung Disease; GERD, gastroesophageal reflux disease.

## References

1. Global Initiative for Chronic Obstructive Lung Disease (GOLD). *Global Strategy for the Diagnosis, Management and Prevention of COPD, 2015*; 2015. Available from: <http://www.goldcopd.org/>. Accessed July 7, 2015.
2. 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.
3. Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax.* 1999;54(7):581–586.

### 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.