

# Use of clinical characteristics to predict spirometric classification of obstructive lung disease

Steven J Pascoe<sup>1</sup>  
Wei Wu<sup>2,3</sup>  
Kathryn A Collison<sup>1</sup>  
Linda M Nelsen<sup>4</sup>  
Keele E Wurst<sup>5</sup>  
Laurie A Lee<sup>6</sup>

<sup>1</sup>Respiratory Medicines Development Center, GSK, Research Triangle Park, NC, USA; <sup>2</sup>Biostatistics, PAREXEL International, Research Triangle Park, NC, USA; <sup>3</sup>Clinical Statistics, GSK, Research Triangle Park, NC, USA; <sup>4</sup>Value Evidence and Outcomes, GSK, Collegeville, PA, USA; <sup>5</sup>Epidemiology, GSK, Collegeville, PA, USA; <sup>6</sup>Research and Development, GSK, Stevenage, UK

**Background:** There is no consensus on how to define patients with symptoms of asthma and chronic obstructive pulmonary disease (COPD). A diagnosis of asthma–COPD overlap (ACO) syndrome has been proposed, but its value is debated. This study (GSK Study 201703 [NCT02302417]) investigated the ability of statistical modeling approaches to define distinct disease groups in patients with obstructive lung disease (OLD) using medical history and spirometric data.

**Methods:** Patients aged  $\geq 18$  years with diagnoses of asthma and/or COPD were categorized into three groups: 1) asthma (nonobstructive; reversible), 2) ACO (obstructive; reversible), and 3) COPD (obstructive; nonreversible). Obstruction was defined as a post-bronchodilator forced expiratory volume in 1 second ( $FEV_1$ )/forced vital capacity  $< 0.7$ , and reversibility as a post-albuterol increase in  $FEV_1 \geq 200$  mL and  $\geq 12\%$ . A primary model (PM), based on patients' responses to a health care practitioner-administered questionnaire, was developed using multinomial logistic regression modeling. Other multivariate statistical analysis models for identifying asthma and COPD as distinct entities were developed and assessed using receiver operating characteristic (ROC) analysis. Partial least squares discriminant analysis (PLS-DA) assessed the degree of overlap between groups.

**Results:** The PM predicted spirometric classifications with modest sensitivity. Other analysis models performed with high discrimination (area under the ROC curve: asthma model, 0.94; COPD model, 0.87). PLS-DA identified distinct phenotypic groups corresponding to asthma and COPD.

**Conclusion:** Within the OLD spectrum, patients with asthma or COPD can be identified as two distinct groups with a high degree of precision. Patients outside these classifications do not constitute a homogeneous group.

**Keywords:** asthma–COPD overlap syndrome, asthma, COPD, differential diagnosis, surveys and questionnaires

## Introduction

Asthma and COPD have classically been considered as distinct diseases with contrasting clinical presentations and underlying pathophysiologies. Asthma is defined as a heterogeneous disease, usually characterized by chronic airway inflammation, with a history of respiratory symptoms (such as wheeze, shortness of breath, chest tightness, or cough) that vary over time and in intensity, and variable expiratory airflow limitation.<sup>1</sup> COPD is defined as a common, preventable, and treatable disease characterized by persistent respiratory symptoms and airflow limitation due to abnormalities of the airway and/or alveolae that are usually caused by significant exposure to noxious particles or gases.<sup>2</sup> However, the current diagnostic criteria for each disease are somewhat qualitative, with several features common to both diseases.<sup>1,2</sup> There has been little objective research

Correspondence: Steven J Pascoe  
Respiratory Medicines Development  
Center, GSK, 5 Moore Drive, Research  
Triangle Park, NC 27709, USA  
Tel +1 919 491 8859  
Email [steven.j.pascoe@gsk.com](mailto:steven.j.pascoe@gsk.com)

into the sensitivity and specificity of diagnostic approaches to differentiate and characterize subjects with OLD.

There has been recent interest in patients showing symptoms of asthma and COPD,<sup>3–6</sup> with diverse proposals for how best to characterize and treat such individuals. It has been postulated that such patients exist in sufficient numbers to identify a new diagnosis of ACOS; this was introduced into the GOLD and GINA guidelines in 2014.<sup>7</sup> However, an alternative perspective exists that ACOS does not represent an independent clinical entity,<sup>7–9</sup> and that it is preferable to classify these patients as ACO. Another approach has focused on defining multiple phenotypes and endotypes within each disease, to achieve a more targeted approach to therapy.<sup>10–12</sup> A further recent approach identifies “treatable traits” (characteristics indicative of a higher likelihood of response to a particular therapy, regardless of disease classification) and largely discards the recognized diagnostic labels.<sup>13</sup>

Due to the lack of consensus on how to define patients with symptoms of asthma and COPD, multiple definitions have been used for these patients. This is reflected in the variation of prevalence estimates<sup>3,6</sup> and has led to the exclusion of these patients from clinical trials for asthma or COPD;<sup>5,14</sup> limited data are therefore available regarding their response to treatment. Clearer diagnostic criteria for these patients are required to allow better trial design and potentially improve treatment strategies.

This study aimed to investigate the ability of algorithmic approaches to define distinct groups of OLD using readily available medical history and spirometry data. The absence of an established “gold standard” test for diagnosing any OLD prevents assessment of new diagnostic tests by standard comparison methods; instead, the level of agreement was assessed between disease classification according to patient-reported characteristics and classification by spirometry.

## Methods

### Study design

This study (GSK Study 201703 [NCT02302417]) included patients from 77 centers across 11 countries (January–May 2015). The full analysis set comprised patients enrolled in the current study as well as patients aged  $\geq 18$  years with clinical diagnoses of asthma and/or COPD who completed the medical history questionnaire and the spirometry assessment as part of their screening for enrollment in two concurrent clinical trials (GSK Studies 200699 [NCT02164539] and 201496 [NCT02299375]) which have since been completed (these patients were included in the full analysis set for this study regardless of their eligibility for inclusion in the other studies). The evaluable population comprised the

subset of patients from the full analysis set who obeyed the spirometric inclusion criteria. The protocol was approved by the Chesapeake Institutional Review Board; all participants provided written informed consent (patient consent provided for participation in GSK Study 200699 [NCT02164539] or 201496 [NCT02299375] allowed the use of questionnaire and spirometry data in the current study). The study protocol is available online from the GSK Clinical Study Register ([https://www.gsk-clinicalstudyregister.com/search/?study\\_ids=201703](https://www.gsk-clinicalstudyregister.com/search/?study_ids=201703)).<sup>15</sup>

### Questionnaire

A 41-item questionnaire, designed in consultation with physicians and the GINA guidelines,<sup>11</sup> recorded patient demographics, symptoms, morbidity, and medical history. It included 38 medical history questions, two questions capturing patients’ perceived diagnoses and the diagnoses of an HCP, and one question for HCPs to identify clinical features that they find helpful for diagnosis. The questionnaire was administered by an HCP during a single designated clinic visit (Study 201703) or during a medical history assessment (Studies 200699 and 201496).

### Spirometry

Data from the following spirometry assessments, performed according to accepted ATS/ERS standards,<sup>16</sup> were recorded in all patients: percentage predicted FEV<sub>1</sub>, pre- and post-bronchodilator FEV<sub>1</sub> (L); pre- and post-bronchodilator FVC (L); pre- and post-bronchodilator FEV<sub>1</sub>/FVC ratio; and reversibility (mL and %); further details are in the Supplementary materials. Obstruction was defined as a post-bronchodilator FEV<sub>1</sub>/FVC ratio  $< 0.7$ , and reversibility was defined as a post-albuterol increase in FEV<sub>1</sub> of  $\geq 200$  mL and  $\geq 12\%$  from pre-bronchodilator values.

### Patients

Eligible participants were aged  $\geq 18$  years with clinical diagnoses of asthma and/or COPD. Patients were classified into one of three spirometric categories: asthma (nonobstructed and reversible), ACO (obstructed and reversible), or COPD (obstructed and nonreversible).

### Statistical analysis methodology

The primary analysis identified a PM utilizing demographics and questionnaire information to predict patient groups corresponding to the spirometrically defined disease groups, based on multinomial logistic regression analysis. Post hoc analysis further identified models to differentiate 1) patients with asthma (MA) and 2) patients with COPD (MC) from

other patients, based on binary logistic regression analysis (details in the Supplementary materials). ROC analysis was performed for both models. Model-based predictions using the full versions of the MA and MC classified patients as having asthma only, COPD only, asthma and COPD, or neither asthma nor COPD. PLS-DA was conducted to further examine the influential predictors and discriminant ability between groups.

Univariate analyses were performed to investigate differences between spirometric cohorts for each questionnaire item; items were classified into types according to the ORs between the ACO cohort and the asthma or COPD cohort (details in the Supplementary materials).

Correlation coefficient analyses examined associations between explanatory variables. Moderately or highly correlated variables were assessed for multicollinearity and excluded as appropriate from the multinomial logistic regression analysis. Multinomial logistic regression modeling was

used to identify discriminatory variables using a split-sample approach (50% of data used to identify predictors; 50% used for performance evaluation) in seven modeling rounds, each with orthogonal partitions of the original sample (Supplementary materials). Discriminatory variables were assessed using model-based predictions relative to spirometry-based classification.

## Results

### Study population

The full analysis set included 1,422 patients; of these, 174 were spirometrically classified as nonobstructed and non-reversible and were excluded. The evaluable population therefore included 1,248 patients classified by spirometry as having asthma (n=307), ACO (n=548), or COPD (n=393).

Some between-group differences were observed in baseline characteristics (Table 1): patients with asthma were significantly younger than those with ACO (age difference

**Table 1** Baseline characteristics and demographics

Characteristics	Asthma, N=307	ACO, N=548	COPD, N=393	Total, N=1,248
Age (years)	42.5 (15.3)	56.5 (12.1)	63.2 (10.7)	55.2 (14.8)
Female, n (%)	201 (65)	256 (47)	172 (44)	629 (50)
Body mass index (kg/m <sup>2</sup> )*	29.36 (7.95)	28.60 (6.06)	28.43 (6.17)	28.74 (6.62)
Smoking status, n (%)*				
Never	255 (83)	211 (39)	46 (12)	512 (41)
Former	38 (12)	193 (36)	180 (46)	411 (33)
Current	14 (5)	139 (26)	165 (42)	318 (26)
Smoking history (pack-years) <sup>†</sup>	10.2 (12.1)	30.8 (26.7)	47.0 (28.8)	37.0 (29.0)
Exacerbation history (past 12 months), n (%) <sup>‡§</sup>				
0 or 1 not leading to hospital admission	249 (81)	432 (80)	259 (66)	940 (76)
≥2 or 1 leading to hospital admission	58 (19)	111 (20)	132 (34)	301 (24)
Pre-bronchodilator FEV <sub>1</sub> (L)	2.325 (0.670)	1.566 (0.514)	1.476 (0.638)	1.724 (0.688)
Post-bronchodilator FEV <sub>1</sub> (L)	2.821 (0.804)	1.987 (0.585)	1.552 (0.657)	2.055 (0.819)
Pre-bronchodilator FEV <sub>1</sub> , % predicted <sup>¶</sup>	72.81 (12.58)	51.74 (14.95)	51.82 (17.57)	56.95 (17.78)
Post-bronchodilator FEV <sub>1</sub> , % predicted <sup>¶</sup>	88.32 (13.96)	65.84 (17.08)	54.61 (18.05)	67.84 (20.94)
Pre-bronchodilator FEV <sub>1</sub> /FVC, %	75.96 (9.08)	51.49 (10.63)	45.15 (21.50)	55.51 (18.95)
Post-bronchodilator FEV <sub>1</sub> /FVC, %	80.71 (6.95)	55.87 (10.32)	45.39 (21.67)	58.68 (19.62)
FEV <sub>1</sub> reversibility (mL)	495.9 (246.5)	421.5 (200.8)	75.8 (103.3)	331.0 (258.4)
FEV <sub>1</sub> reversibility (%)	22.06 (10.72)	29.28 (17.14)	5.97 (7.23)	20.16 (16.56)
Inhaled maintenance therapy taken in the past 6 months, n (%) <sup>**</sup>				
None	62 (20)	59 (11)	76 (19)	197 (16)
ICS	76 (25)	63 (11)	23 (6)	162 (13)
LABA	3 (<1)	3 (<1)	2 (<1)	8 (<1)
LAMA	1 (<1)	21 (4)	41 (10)	63 (5)
ICS/LABA	164 (53)	348 (64)	133 (34)	645 (52)
ICS/LAMA	0	3 (<1)	5 (1)	8 (<1)
LAMA/LABA	0	6 (1)	21 (5)	27 (2)
ICS/LABA/LAMA	1 (<1)	45 (8)	92 (23)	138 (11)

**Notes:** Values are reported as mean (standard deviation) unless otherwise stated. \*Asthma, n=307; ACO, n=543; COPD, n=391; total, n=1,241. †Asthma, n=52; ACO, n=331; COPD, n=345; total, n=728. ‡Pack-years = (number of cigarettes smoked per day/20) × number of years smoked. §Exacerbations not leading to hospitalization were defined as episodes requiring treatment with oral/systemic corticosteroids and/or antibiotics. ¶Asthma, n=307; ACO, n=548; COPD, n=392; total, n=1,247. \*\*Responses were mutually exclusive.

**Abbreviations:** ACO, asthma–COPD overlap; COPD, chronic obstructive pulmonary disease; FEV<sub>1</sub>, forced expiratory volume in 1 second; FVC, forced vital capacity; ICS, inhaled corticosteroid; LABA, long-acting β<sub>2</sub>-agonist; LAMA, long-acting muscarinic antagonist.

[95% CI]: -14.0 years [-15.8, -12.2];  $p < 0.001$ ), while patients with COPD were significantly older than those with ACO (age difference [95% CI]: 6.7 years [5.0, 8.3];  $p < 0.001$ ). A higher proportion of patients with asthma were female compared with ACO or COPD, and fewer patients with asthma had a history of smoking. The proportion of patients experiencing exacerbations was highest in the COPD cohort, while baseline lung function was generally poorer in patients with COPD, but higher in patients with asthma, than in patients with ACO (Table 1).

### Univariate analysis of questionnaire items

Three questionnaire items were classified as Type 1; each of these had ORs of COPD and asthma to ACO of  $< 1$  (Table 2). Five items were classified as Type 2, with ACO between asthma and COPD in terms of likelihood (for these items, ORs of asthma to ACO were  $> 1$  and ORs of COPD to ACO were  $< 1$ ). Ten items were classified as Type 3a (with similarity to asthma and difference from COPD), while 13 items were classified as Type 3c (with similarity to COPD and difference from asthma). Nine items were classified as Type 4, with no significant difference between asthma and ACO or COPD and ACO.

### Correlation analysis

Multiple variables were found to be related with a correlation coefficient magnitude  $> 0.5$  (Table 3). The variables carried forward for the multinomial logistic regression analysis were age, sex, smoking status (never/current/past), BMI, exacerbation history (0 or 1 not leading to hospital admission/ $\geq 2$  or 1 leading to hospital admission), and questionnaire items 1–36 (excluding items 5, 7, 13, 22, and 34).

### Multinomial logistic regression

Results of the model-based predictions in each round of the multinomial logistic regression analysis are summarized in Table 4 for variables that were significant in at least four of the seven rounds. These items were included in the PM (Table 5) along with age, sex, smoking status, BMI, and exacerbation history.

Concordance between disease classifications by spirometry and model-based prediction was generally consistent throughout the seven rounds of cross-validation (data not shown). In these rounds, the model-predicted diagnosis of asthma in patients with spirometry-classified asthma ranged from 54% to 69%, the predicted diagnosis of COPD in patients with spirometry-classified COPD ranged from 50% to 59%, and the predicted diagnosis of ACO in patients

with spirometry-classified ACO ranged from 52% to 61%. Where the model-predicted diagnosis was not consistent with the spirometric classification, patients with spirometry-classified asthma were more commonly misdiagnosed with ACO (26%–39% of patients across all modeling rounds) than with COPD (3%–11%), and patients with spirometry-classified COPD were more commonly misdiagnosed with ACO (37%–45%) than with asthma (2%–7%).

The level of agreement between disease classifications predicted by the PM and by spirometry is shown in Table 6. The accordance of model-based prediction with spirometric disease classification was 69% for asthma, 70% for ACO, and 64% for COPD, with an overall accuracy of 68%. The discordance between asthma and COPD was low: 3% of patients with spirometry-classified asthma had a model-based diagnosis of COPD, and 3% of patients with spirometry-classified COPD had a model-based diagnosis of asthma. For classification of patients with or without ACO, the PM performed with 70% sensitivity and 70% specificity, with a PPV of 64% and an NPV of 75%. The estimated Kappa coefficient was 0.65, with an estimation precision of 0.07 (width of the 95% CIs).

### Post hoc analysis

Three versions of the models for asthma and COPD were identified: a basic model (including age, sex, smoking status, and either BMI [MA] or exacerbation history [MC] as explanatory variables), a reduced model (comprising the basic model characteristics plus those found to be significant predictors of disease in at least four out of seven rounds of logistic regression modeling), and a full model (comprising the basic model characteristics plus responses to questionnaire items 1–36, excluding item 30). ROC analysis demonstrated that all versions of each model had high discriminative power to distinguish between patients with and without the disease of interest. For the asthma model, the AUC was 0.9376 for the full model (Figure 1A) and slightly lower for the basic and reduced models (0.8742 and 0.9097, respectively). The full version of the MA performed with 95% specificity and 73% sensitivity, with a PPV of 0.82 and an NPV of 0.91. For the COPD model, the AUC was 0.8693 for the full model (Figure 1B) and slightly lower for the basic and reduced models (0.8114 and 0.8209, respectively). The full version of the MC performed with 89% specificity and 62% sensitivity, with a PPV of 0.72 and an NPV of 0.83.

The distributions of patients according to the presence or absence of asthma and COPD, as predicted by the full models for asthma and COPD, in each spirometric cohort

**Table 2** Results of the univariate analysis

Type*	Item	Question	Odds ratio (asthma to ACO)	p-value	Odds ratio (COPD to ACO)	p-value
1	2	How often does the subject need to keep a rescue inhaler with him/her?	0.33	0.004	0.19	<0.001
1	18	If the subject stops taking his/her regular (maintenance/controller) respiratory medications, how do his/her respiratory symptoms change?	0.31	0.002	0.31	0.002
1	21	On average how frequently does the subject use his/her rescue medication?	0.70	0.050	0.61	0.002
2	4	In the past two (2) years has the subject had periods of several weeks or longer where he/she had no respiratory symptoms?	2.18	<0.001	0.58	<0.001
2	5	On the days the subject does not have daytime respiratory symptoms, how often does the subject wake at night due to respiratory symptoms?	1.47	0.019	0.60	<0.001
2	7	Does the subject tend to have both "good" and "bad" days with regard to breathing, or are most days about the same?	1.55	0.005	0.69	0.006
2	8	On average, how much do the subject's good and bad days differ?	1.66	0.004	0.62	<0.001
2	11	Has the subject ever had nasal allergies or eczema?	3.31	<0.001	0.62	<0.001
3a	1	How well does the subject's quick relief inhaler (rescue inhaler) provide symptom relief?	1.40	0.689	0.25	0.008
3a	3	How does the subject describe his/her respiratory disease over the past two (2) years?	1.36	0.110	0.71	0.028
3a	6	Typically, how are the subject's respiratory symptoms during the night and upon awakening compared to respiratory symptoms during the day?	0.90	0.724	0.58	0.035
3a	14	Do the subject's respiratory symptoms get worse after exposure to cold air or weather changes?	1.35	0.127	0.71	0.036
3a	15	Do the subject's respiratory symptoms get worse with exposure to air pollution or noxious fumes?	1.12	0.573	0.69	0.025
3a	20	How often does the subject typically have respiratory symptoms at night?	1.20	0.241	0.75	0.033
3a	22	How often do the subject's respiratory symptoms disturb his/her sleep?	1.39	0.081	0.59	<0.001
3a	31	In the past 12 months, how frequently has the subject experienced chest tightness?	1.20	0.212	0.77	0.046
3a	33c	Cough is the most bothersome symptom	1.03	0.878	1.65	0.001
3a	33w	Wheezing is the most bothersome symptom	1.20	0.396	0.58	0.021
3c	10	When the subject's respiratory symptoms are bad, how long until they return to normal?	1.34	0.039	1.18	0.219
3c	12	How many members of the subject's immediate biological family have had asthma, nasal allergies or eczema?	3.24	<0.001	0.96	0.729
3c	13	Does the subject's respiratory symptom get worse after exposure to pollen or pets?	2.95	<0.001	0.79	0.077
3c	16	Does the subject react emotionally to distress (eg, cry easily during a sad film)?	1.36	0.033	0.94	0.619
3c	19	How many days in a week does the subject typically have respiratory symptoms during the day?	0.59	0.006	0.82	0.309
3c	24	How often would the subject describe themselves as feeling anxious?	1.60	0.030	1.04	0.818
3c	26	Do colds often go to the subject's chest?	5.01	0.003	0.80	0.380
3c	27	Over the past two years, how often has the subject had periods of frequent cough that lasted for several days or more?	2.91	0.003	0.84	0.449
3c	29	In the past 12 months, how frequently has the subject experienced breathlessness?	0.56	<0.001	0.88	0.447
3c	30	In the past 12 months, how frequently has the subject coughed up sputum (phlegm or mucus)?	0.61	<0.001	1.28	0.097
3c	34	When the subject is at his/her best, how much exercise can he/she do before the subject gets breathless?	0.40	<0.001	1.14	0.485
3c	33b	Breathlessness is the most bothersome symptom	0.66	0.004	0.84	0.197
3c	33t	Chest tightness is the most bothersome symptom	2.68	<0.001	1.01	0.960
4	9	Describe how quickly a good day may change to a bad day due to respiratory symptoms.	0.89	0.408	1.18	0.220
4	17	How much of an impact does emotional distress have on the subject's respiratory symptoms?	1.20	0.221	0.92	0.528
4	23	How much impact do the subject's respiratory symptoms have on his/her energy level?	0.64	0.141	0.84	0.556
4	25	When the subject coughs, how often does he/she bring up phlegm or mucus?	0.62	0.112	0.70	0.227
4	28	In the past 12 months, how frequently has the subject experienced cough?	0.81	0.186	1.35	0.063
4	32	In the past 12 months, how frequently has the subject experienced wheezing?	1.17	0.319	0.98	0.887
4	33s	Sputum is the most bothersome symptom	0.86	0.624	0.92	0.781
4	35	How often would the subject describe themselves as feeling depressed?	1.25	0.138	1.15	0.316
4	36	How scared or worried was the subject about his/her lung function?	0.87	0.470	0.77	0.135

**Notes:** \*Type 1: ORs of the COPD and asthma cohorts to the ACO cohort both significantly different from 1 in the same direction; Type 2: ORs both significantly different from 1 in different directions; Type 3: ORs significantly different from 1 for one comparison only (Type 3a: ACO was similar to asthma but different from COPD; Type 3c: ACO was similar to COPD but different from asthma); Type 4: ORs showed no significant difference between either asthma and ACO or COPD and ACO. The reference group for the univariate analysis was represented by the first answer for each question.

**Abbreviations:** ACO, asthma-COPD overlap; COPD, chronic obstructive pulmonary disease; OR, odds ratio.

**Table 3** Variables with correlation coefficient magnitude >0.5

Type	Variables	Items	n	Correlation coefficient	p-value	
Spearman	Nighttime symptoms and sleep	5*, 20	1,245	0.57	<0.001	
		5*, 22*	1,245	0.66	<0.001	
		20, 22*	1,244	0.68	<0.001	
	Good and bad days	7*, 8	1,241	-0.70	<0.001	
		Allergy, eczema	11, 13*	1,240	0.51	<0.001
		Daytime symptoms, breathlessness	19, 29	1,248	0.59	<0.001
		Cough/sputum	25, 28	1,248	0.53	<0.001
			25, 30	1,247	0.69	<0.001
			27, 28	1,246	0.51	<0.001
			28, 30	1,247	0.66	<0.001
Pearson	Age; age first treated with an inhaler	-	1,234	0.75	<0.001	
	Age; age first treated with an ICS inhaler	-	1,167	0.82	<0.001	
	Age first treated with an inhaler; age first treated with an ICS inhaler	-	1,160	0.87	<0.001	

**Notes:** \*Item excluded from the multinomial logistic regression analysis. “-”, not applicable.

**Abbreviation:** ICS, inhaled corticosteroid.

are presented in Figure 2. The majority of patients in the asthma spirometric cohort (208/287, 72%) were classified by the models as having asthma and not COPD; similarly, the majority of patients in the COPD spirometric cohort (229/369, 62%) were classified as having COPD and not asthma. A total of 11 out of 1,161 patients (1%) were classified by the models as having asthma and not COPD but were spirometrically classified as having COPD, or vice versa. Most patients (384/505, 76%) in the ACO spirometric cohort were classified by the models as having neither asthma nor COPD, while only 3 out of 1,161 patients (<1%) in the entire population were classified as having asthma and COPD.

PLS-DA using data on age, sex, smoking status, BMI, exacerbation incidence, and responses to questionnaire items 1–36 identified reasonable separation between the groups of patients identified as having asthma and as having COPD by spirometry, with a  $Q^2$  value of 0.59 (Figure 3). The most influential predictors for distinguishing asthma and COPD, in order of decreasing Variance Important in Projection measure, were smoking status, age, questionnaire item 11 (presence of nasal allergies/eczema), questionnaire item 30 (frequency of coughing up sputum), and questionnaire item 4 (presence of periods of several weeks or longer with no respiratory symptoms). When PLS-DA using the same input variables was applied to assess the separation between all

**Table 4** Summary of model-based prediction: significant predictors in at least four out of seven rounds of modeling

Effect	Modeling round, p-value							Number of rounds in which item was found to be a significant predictor
	1	2	3	4	5	6	7	
Age	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	7
Sex	<0.001	0.577	0.045	0.116	<0.001	0.004	0.057	6
Smoking status	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	7
Exacerbation history	0.006	0.002	<0.001	<0.001	0.004	<0.001	0.011	7
Item 8	0.027	0.002	0.002	0.021	0.014	<0.001	<0.001	7
Item 2	0.052	0.005	0.002	0.008	<0.001		0.025	6
Item 1	0.047	0.038		0.084	0.079	0.016		5
Item 10	0.012	0.046	0.012	0.090	0.021			5
Item 12			0.007	<0.001	0.002	0.023	<0.001	5
Item 18	0.004	0.028	0.003	0.002		0.013		5
Item 29		0.006	0.017	0.015		0.007	0.023	5
Item 4			0.005	0.008	0.029		0.117	4
Item 35		0.043	0.004	0.004			0.042	4
BMI	0.013	0.030	0.012			0.017		4

**Note:** Please refer to Table 2 for references to item numbers.

**Abbreviation:** BMI, body mass index.

**Table 5** Questionnaire items included in the final model

Item	Question
1	How well does the subject's quick relief inhaler (rescue inhaler) provide symptom relief?
2	How often does the subject need to keep a rescue inhaler with him/her?
4	In the past two years has the subject had periods of several weeks or longer where (s)he had no respiratory symptoms?
8	On average, how much do the subject's good and bad days differ?
10	When the subject's respiratory symptoms are bad, how long until they return to normal?
12	How many members of the subject's immediate biological family have had asthma, nasal allergies or eczema?
18	If the subject stops taking his/her regular (maintenance/controller) respiratory medications, how do his/her respiratory symptoms change?
29	In the past 12 months, how frequently has the subject experienced breathlessness?
35	How often would the subject describe themselves as feeling depressed?

**Note:** Age, sex, smoking status, body mass index, and exacerbation history were also included in the final model.

three cohorts identified by spirometry, greater homogeneity was observed (Figure 4) with a  $Q^2$  value of 0.18. When lung function data (percentage reversibility at screening, post-bronchodilator  $FEV_1/FVC$  ratio, and pre-bronchodilator  $FEV_1$ ) were added to the inputs to the PLS-DA, increased discrimination was observed between the three spirometrically identified cohorts with a  $Q^2$  of 0.41 (Figure 5).

## Discussion

This study aimed to investigate the use of statistical modeling approaches to define distinct disease groups within the spectrum of OLD. A secondary aim was to investigate whether patients with features of asthma and COPD could be divided into distinct groups more closely resembling those with asthma or COPD, or whether these patients presented mixed characteristics, suggestive of an intermediate phenotype.

**Table 6** Summary of model-based prediction relative to disease classification by spirometry

Outcome predicted by clinical features based on PM	Disease classification by spirometry			
	Asthma	ACO	COPD	Total
Asthma				
n	<b>208</b>	66	12 <sup>#</sup>	286
Column percent	<b>69%</b>	12%	3% <sup>#</sup>	24%
ACO				
n	82	<b>369</b>	124	575
Column percent	27%	<b>70%</b>	32%	47%
COPD				
n	10 <sup>#</sup>	95	<b>246</b>	351
Column percent	3% <sup>#</sup>	18%	<b>64%</b>	29%
Total				
n	300	530	382	1,212
Kappa coefficient (95% CI)	0.65 (0.62, 0.69)			

**Notes:** Frequencies on the diagonal (in bold typeface) indicate agreement between classification by spirometry and by predictive model. Frequencies off the diagonal indicate misclassifications. These are considered more allowable between adjacent categories (numbers in plain text) than between nonadjacent categories (numbers with #).

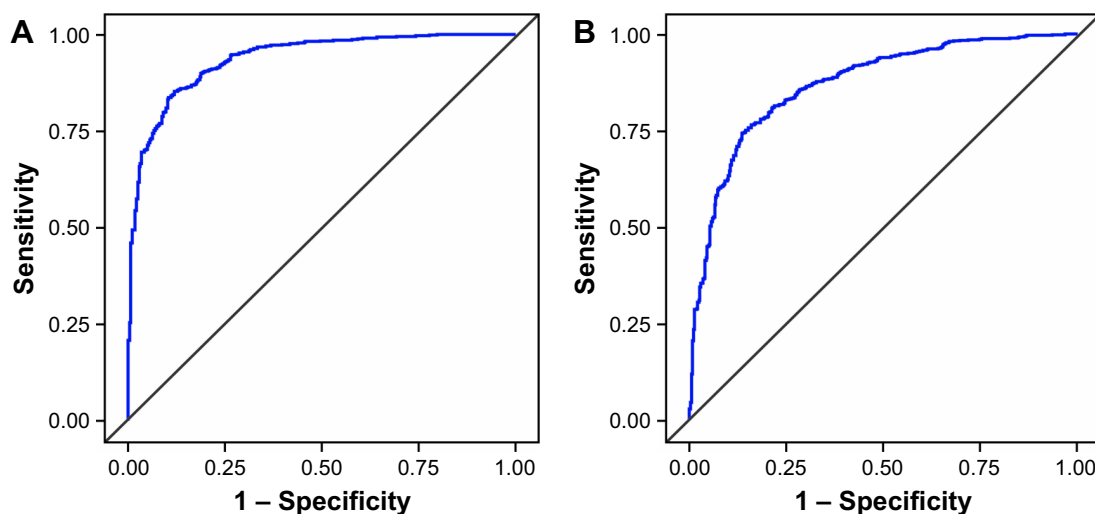
**Abbreviations:** ACO, asthma–COPD overlap; CI, confidence interval; COPD, chronic obstructive pulmonary disease; PM, primary model.

Patient characteristics and questionnaire responses of the ACO population did not align consistently with either the asthma or the COPD populations, with several measures aligned midway between the asthma and COPD populations and some measures aligned more closely with either the asthma or the COPD population. Patient characteristics at baseline, particularly the lower age and greater likelihood of being female of patients with asthma, and higher age of patients with COPD, compared with patients with ACO, were generally consistent with observations from previous studies.<sup>17–21</sup>

In the primary analysis, many individual questionnaire items had high degrees of statistical significance for the differences between the disease groups, but the magnitudes of the differences were small and responses were heterogeneous across the populations. This indicates that any response alone would have low discriminatory power between the disease classifications; combining the responses enables more accurate discrimination.

The PM developed here predicted spirometric classifications based on patients' clinical histories with modest sensitivity. A model classification of asthma was assigned rarely to patients with spirometrically classified COPD (4% of those with this model classification), more often to patients with spirometrically classified ACO (23%), and frequently to those with spirometrically classified asthma (73%). Similarly, a model classification of COPD was assigned rarely to patients with spirometrically classified asthma (3% of those with this model classification), more often to patients with spirometrically classified ACO (27%), and frequently to those with spirometrically classified COPD (70%). These data support the concept that asthma and COPD are independent phenotypes.

To explore this further, models were developed to discriminate between patients with or without asthma (MA) and patients with or without COPD (MC). The full versions of



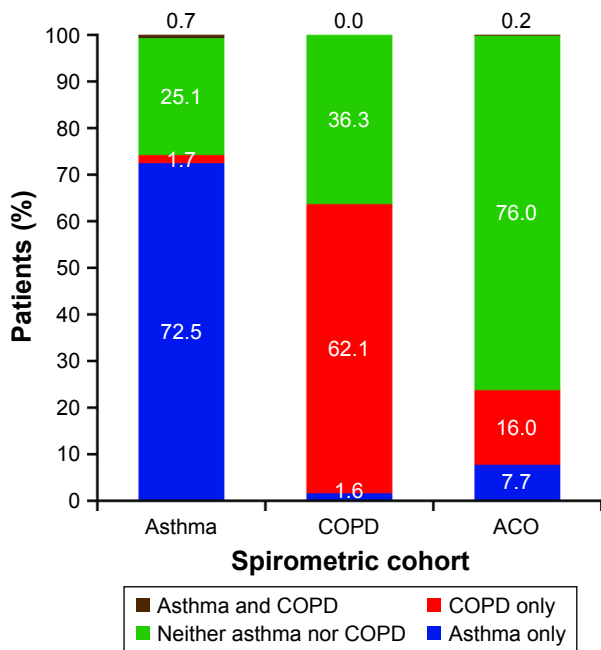
**Figure 1** ROC curves for the full independent models: (A) independent model for asthma; (B) independent model for COPD. **Abbreviations:** COPD, chronic obstructive pulmonary disease; ROC, receiver operating characteristic.

both models were highly discriminative. Using these models, very few patients (0.3%) were classified as having both asthma and COPD, and few patients (0.9%) were classified as having asthma alone but were spirometrically classified as having COPD, or vice versa. A considerable proportion of patients in each spirometric group were classified as having neither asthma nor COPD.

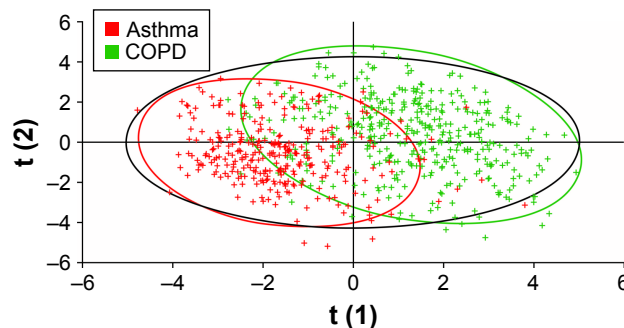
PLS-DA was used to further investigate the relationship between disease groups, with  $Q^2$  (which varies from negative

to +1 with higher positive values representing better separation) providing a measure of discriminatory ability. A high degree of separation was observed between the groups identified by the PM as having asthma and COPD, reflecting a high discriminatory ability of the PM for these two conditions. However, when the ACO group was added, considerable overlap was seen between the groups. The degree of overlap was reduced somewhat by adding patients' spirometric data to the model.

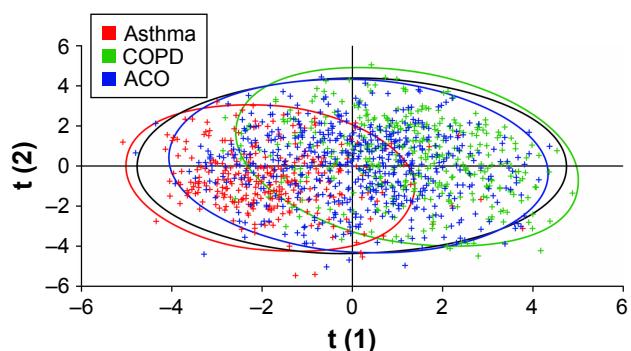
Taken together, these results show that a positive diagnosis of asthma (with 95% specificity) and COPD (89% specificity) can be made with a high degree of confidence that the diagnosis can discriminate between these two distinct phenotypic groups. However, the group of patients with features of both asthma and COPD was not homogeneous and had considerable phenotypic overlap with both asthma and COPD. While it was not the primary aim of this study to develop questionnaires for clinical use, these results



**Figure 2** Distribution of patients in each spirometric cohort by asthma and COPD status according to the full models for asthma and COPD. **Abbreviations:** ACO, asthma-COPD overlap; COPD, chronic obstructive pulmonary disease.



**Figure 3** Partial least squares discriminant analysis of the asthma and COPD cohorts defined by spirometry. **Note:** Data points are color-coded according to spirometric classification. **Abbreviation:** COPD, chronic obstructive pulmonary disease.



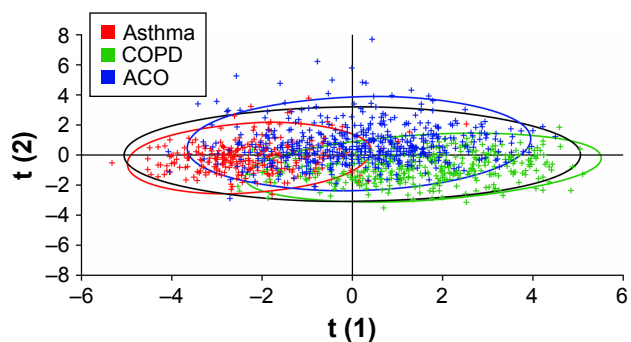
**Figure 4** Partial least squares discriminant analysis of the asthma, COPD, and ACO cohorts defined by spirometry.

**Note:** Data points are color-coded according to spirometric classification.

**Abbreviations:** ACO, asthma–COPD overlap; COPD, chronic obstructive pulmonary disease.

demonstrate the feasibility of using statistical modeling approaches to discriminate between distinct disease groups in the spectrum of OLD, as well as the independence of the asthma and COPD phenotypes. The data obtained can inform clinical research into how best to treat patients with symptoms of both asthma and COPD, as they support the approach of identifying “treatable traits”, based on the individual characteristics of the patient, rather than the use of the diagnosis of ACOS. Additionally, particular clinical features, including age, sex, smoking status, BMI, exacerbation history, response to rescue medication, symptomatic periodicity, and history of allergy or atopy, have been identified as important in discriminating between patients with different types of OLD; these criteria could perhaps be employed in addition to spirometric assessment in clinical practice.

As most clinical studies in asthma and COPD have excluded patients with OLD without a clearly defined diagnosis of either condition, their results may not be relevant to these patients. This omission seems especially important in



**Figure 5** Partial least squares discriminant analysis of the asthma, COPD, and ACO cohorts defined by spirometry after addition of spirometric data to questionnaire responses.

**Note:** Data points are color-coded according to spirometric classification.

**Abbreviations:** ACO, asthma–COPD overlap; COPD, chronic obstructive pulmonary disease.

the context of diverging guidelines and treatment options for asthma and COPD (eg, the early and widespread use of ICSs in asthma,<sup>1</sup> contrasting with the preference for long-acting bronchodilators in COPD with ICSs reserved for patients suffering exacerbations on initial therapy<sup>2</sup>). The advent of monoclonal antibody therapies for asthma<sup>22</sup> is likely to increase the differential treatment of OLD. We believe that there is an urgent need for studies examining patients who are not clearly defined as having “pure” asthma or “pure” COPD, with a corresponding recognition of this area of interest by journals and regulators. In the absence of prospective data, it seems advisable that clinicians emphasize identification of patients with asthma or COPD and prescribe treatments based on extensive clinical trial data. For patients without a positive diagnosis of asthma or COPD, it seems advisable for clinicians to adopt a treatable traits approach, based on the individual characteristics of the patients.

While some questionnaires have been developed to diagnose either asthma or COPD,<sup>23–25</sup> few studies have investigated the use of patient questionnaires for diagnosis of ACO. Consistent with our findings that asthma can be clearly distinguished from COPD, some other questionnaires have achieved this distinction: one four-item questionnaire successfully distinguished patients with asthma and COPD in accordance with GOLD and GINA guidelines, reporting a sensitivity of 87.6% and a specificity of 87.2% for diagnosis of COPD; ~20% of the population examined had no clear distinction between asthma and COPD, potentially identifying patients with ACO.<sup>26</sup> Another nine-item differential diagnosis questionnaire reported a sensitivity of 72.0% and a specificity of 82.7% for COPD diagnosis in patients aged  $\geq 40$  years with prior diagnosis of OLD.<sup>27</sup> Both of these questionnaires included age of onset, smoking history, and features of a patient’s cough as predictors of diagnosis; these items were also included in the questionnaire developed here. More recently, a scoring system was proposed based on three questions (relating to age of onset and nature of breathlessness) and a chest X-ray; this also successfully identified patients with asthma and with COPD, with some patients (25%–26% in the validation dataset) remaining difficult to differentiate.<sup>28</sup> However, classification of potential ACO was not attempted.

In the current study, the spirometric classifications of asthma, COPD, and ACO were based on the presence or absence of obstruction and reversibility. Obstruction was defined as a post-bronchodilator FEV<sub>1</sub>/FVC ratio  $< 0.7$  in accordance with the GOLD guidelines;<sup>2,11</sup> however, it has been proposed by the ATS/ERS task force that a definition

of an FEV<sub>1</sub>/vital capacity ratio below the lower limit of the normal distribution might provide greater accuracy.<sup>29</sup> Additionally, the definition of reversibility proposed by the ATS/ERS task force includes a post-albuterol increase of  $\geq 200$  mL and  $\geq 12\%$  from pre-bronchodilator values in FVC as an alternative to FEV<sub>1</sub>, while this study only employed FEV<sub>1</sub>. Further research is required to determine which criteria would enable optimal diagnostic accuracy; however, there is also a need to balance this with the practicalities of routine clinical use.

It should be noted that the spirometric groups defined as asthma, COPD, and ACO in this study cannot be considered to represent groups with those diagnoses exactly, as clinical diagnoses are based on additional features to spirometric measurements. This could be considered as one of the limitations of this study, as for example, a diagnosis of COPD is often based on a history of tobacco or biomass exposure. However, by excluding non-spirometric criteria from the disease classifications in this study, it was possible to define phenotypic clusters based on these characteristics and assess the extent to which these clusters overlap; ultimately, this enabled us to establish whether ACO can be considered to be a single independent phenotype. For example, by excluding smoking history from the disease definitions, it was possible to consider whether patients with spirometric characteristics typical of COPD and no smoking history have similar phenotypic characteristics to patients with smoking-related COPD, and similarly whether patients with spirometric characteristics typical of asthma but who are smokers have similar phenotypic characteristics to patients with asthma and no smoking history. In fact, data from the questionnaire demonstrated that tobacco exposure was considered as part of the physician-assigned COPD diagnosis in the majority of patients with spirometrically classified COPD: item 41 of the questionnaire asked physicians to choose any number of applicable responses from a list to answer the question: “What clinical features most helped your diagnosis?” The response “tobacco smoking history” was selected by physicians for 83% of patients with a spirometric classification of COPD (corresponding values for patients with spirometric classifications of asthma and ACO were 13% and 51%, respectively). While there is no consensus on the diagnosis criteria for ACO, the criterion of obstruction included in this study (based on the GINA and GOLD guidance on ACO published in 2014<sup>11</sup>) has been included in more recently proposed criteria for a diagnosis of ACO.<sup>30,31</sup> However, recent publications suggest that the criterion for reversibility used in this study ( $\geq 200$  mL and  $\geq 12\%$ ) may indicate reversible COPD rather than ACO, with higher reversibility ( $\geq 400$  mL)

being more suggestive of ACO.<sup>32</sup> With the ongoing research into biomarkers and the use of computed tomography to characterize lung diseases,<sup>32</sup> more conclusive classifications may be achievable.

A limitation of this study is that the numbers of patients in each disease group were based on preestablished targets to enable statistical comparisons between groups, and therefore are not indicative of the prevalence of each disease. Patients without OLD were not included, so the ability of the questionnaire to distinguish patients with or without OLD could not be assessed. No long-term outcome data were collected, so it is not possible to make an assessment of any difference in prognoses between the groups identified; this would require extensive longitudinal data capture and would be an interesting area for future study. An additional limitation was that questionnaire responses were taken at a single time point; no information is available on the stability of responses over time. Questionnaires can be context specific and may not be relevant to all countries, practice settings and uses, or all patient groups. However, as previously stated, it was not the main intent of this study to develop a questionnaire for use in clinical practice, since the use of complex algorithms, while highly informative from a research perspective, does not easily translate into a clinical setting.

In conclusion, this study shows that a model derived from responses to medical history questions typically employed in clinical practice is able to discriminate between spirometric classifications of asthma, COPD, and ACO with a modest degree of sensitivity. The most powerful discriminatory factors were age, sex, smoking status, BMI, exacerbation history and responses to questions about response to treatment, symptomatic periodicity, and history of allergy or atopy. Two distinct phenotypic groups could be defined corresponding to spirometric diagnoses of asthma and COPD. Patients classified by the models as having neither asthma nor COPD were most commonly classified by spirometry as having potential ACO and formed the majority of patients in this cohort but represented minorities in the spirometric asthma and COPD cohorts. The distribution of the PM-defined phenotypes after incorporating lung function into the model strongly suggests that patients with OLD are best classified into two clearly defined phenotypes of asthma and COPD, with a significant proportion having intermediate phenotypes. This suggests that OLD is phenotypically best represented as a spectrum with clearly distinct poles.

## Abbreviations

ACO, asthma–COPD overlap; ACOS, asthma–COPD overlap syndrome; ATS, American Thoracic Society; AUC, area

under the ROC curve; BMI, body mass index; CI, confidence interval; COPD, chronic obstructive pulmonary disease; ERS, European Respiratory Society; FEV<sub>1</sub>, forced expiratory volume in 1 second; FVC, forced vital capacity; GINA, Global Initiative for Asthma; GOLD, Global Initiative for Chronic Obstructive Lung Disease; HCP, health care practitioner; ICS, inhaled corticosteroid; MA, model for asthma; MC, model for COPD; NPV, negative predictive value; OLD, obstructive lung disease; OR, odds ratio; PLS-DA, partial least squares discriminant analysis; PM, primary model; PPV, positive predictive value; ROC, receiver operating characteristic.

## Acknowledgments

This study was funded by GSK (study number: 201703). The funder had a role in study design, data analysis, data interpretation, and writing of the report. The corresponding author had full access to all the data and the final responsibility to submit for publication. Editorial assistance in the preparation of this manuscript (in the form of writing assistance, including developing of the initial draft based on author direction, assembling tables and figures, collating and incorporating authors' comments, grammatical editing, and referencing) was provided by Elizabeth Jameson, PhD, and Natasha Thomas, PhD, of Fishawack Indicia Ltd, UK, and was funded by GSK.

## Author contributions

WW contributed to the study concept and design and was involved in data acquisition, analysis, and interpretation. SJP, KAC, LMN, KEW, and LAL contributed to the study concept and design and were involved in data analysis and interpretation. All authors were involved in preparation and review of the manuscript and approved the final version to be submitted.

## Disclosure

SJP, KAC, LMN, KEW, and LAL are employees of GSK and hold stocks/shares in GSK. WW was an employee of GSK at the start of the study, owns GSK stock, and moved to PAREXEL during the study conduct, where she is currently an employee. The authors report no other conflicts of interest in this work.

## References

- Global Initiative for Asthma. Global strategy for asthma management and prevention [updated 2017]. Available from: <http://ginasthma.org/2017-gina-report-global-strategy-for-asthma-management-and-prevention/>. Accessed November 28, 2017.
- Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease [updated 2017]. Available from: <http://goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/>. Accessed February 9, 2018.
- Alshabanat A, Zafari Z, Albanyan O, Dairi M, FitzGerald JM. Asthma and COPD overlap syndrome (ACOS): a systematic review and meta analysis. *PLoS One*. 2015;10(9):e0136065.
- Lenz PH, Panos RJ. Asthma and COPD – overlapping disorders or distinct processes? In: Panos RJ, editor. *COPD Clinical Perspectives*. InTech; Rijeka, Croatia. 2014.
- van den Berge M, Aalbers R. The asthma-COPD overlap syndrome: how is it defined and what are its clinical implications? *J Asthma Allergy*. 2016;9:27–35.
- Wurst KE, Kelly-ReifK, Bushnell GA, Pascoe S, Barnes N. Understanding asthma-chronic obstructive pulmonary disease overlap syndrome. *Respir Med*. 2016;110:1–11.
- Cazzola M, Rogliani P. Do we really need asthma-chronic obstructive pulmonary disease overlap syndrome? *J Allergy Clin Immunol*. 2016;138(4):977–983.
- Kostikas K, Clemens A, Patalano F. The asthma-COPD overlap syndrome: do we really need another syndrome in the already complex matrix of airway disease? *Int J Chron Obstruct Pulmon Dis*. 2016;11:1297–1306.
- Rogliani P, Ora J, Puxeddu E, Cazzola M. Airflow obstruction: is it asthma or is it COPD? *Int J Chron Obstruct Pulmon Dis*. 2016;11:3007–3013.
- Segreti A, Stirpe E, Rogliani P, Cazzola M. Defining phenotypes in COPD: an aid to personalized healthcare. *Mol Diagn Ther*. 2014;18(4):381–388.
- Skloot GS. Asthma phenotypes and endotypes: a personalized approach to treatment. *Curr Opin Pulm Med*. 2016;22(1):3–9.
- Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med*. 2012;18(5):716–725.
- Agusti A, Bel E, Thomas M, et al. Treatable traits: toward precision medicine of chronic airway diseases. *Eur Respir J*. 2016;47(2):410–419.
- Gibson PG, Simpson JL. The overlap syndrome of asthma and COPD: what are its features and how important is it? *Thorax*. 2009;64(8):728–735.
- GlaxoSmithKline. The utility of a clinical questionnaire to identify subjects with features of both asthma and COPD. 2016. Available from: [https://www.gsk-clinicalstudyregister.com/search/?study\\_ids=201703](https://www.gsk-clinicalstudyregister.com/search/?study_ids=201703). Accessed June 9, 2016.
- Miller MR, Hankinson J, Brusasco V, et al; ATS/ERS Task Force. Standardisation of spirometry. *Eur Respir J*. 2005;26(2):319–338.
- Blanchette CM, Gutierrez B, Ory C, Chang E, Akazawa M. Economic burden in direct costs of concomitant chronic obstructive pulmonary disease and asthma in a Medicare Advantage population. *J Manag Care Pharm*. 2008;14(2):176–185.
- 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.
- Kauppi P, Kupiainen H, Lindqvist A, et al. Overlap syndrome of asthma and COPD predicts low quality of life. *J Asthma*. 2011;48(3):279–285.
- Kumbhare S, Pleasants R, Ohar JA, Strange C. Characteristics and prevalence of asthma/chronic obstructive pulmonary disease overlap in the United States. *Ann Am Thorac Soc*. 2016;13(6):803–810.
- Menezes AM, Montes de Oca M, Perez-Padilla R, et al; PLATINO Team. Increased risk of exacerbation and hospitalization in subjects with an overlap phenotype: COPD-asthma. *Chest*. 2014;145(2):297–304.
- Mitchell PD, El-Gammal AI, O'Byrne PM. Emerging monoclonal antibodies as targeted innovative therapeutic approaches to asthma. *Clin Pharmacol Ther*. 2016;99(1):38–48.
- Levy ML, Fletcher M, Price DB, Hausen T, Halbert RJ, Yawn BP. International Primary Care Respiratory Group (IPCRG) guidelines: diagnosis of respiratory diseases in primary care. *Prim Care Respir J*. 2006;15(1):20–34.
- van Schayck CP, Halbert RJ, Nordyke RJ, Isonaka S, Maroni J, Nonikov D. Comparison of existing symptom-based questionnaires for identifying COPD in the general practice setting. *Respirology*. 2005;10(3):323–333.

25. Martinez FJ, Mannino D, Leidy NK, et al; High-Risk-COPD Screening Study Group. A new approach for identifying patients with undiagnosed chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2017;195(6):748–756.
26. Beeh KM, Kornmann O, Beier J, Ksoll M, Buhl R. Clinical application of a simple questionnaire for the differentiation of asthma and chronic obstructive pulmonary disease. *Respir Med*. 2004;98(7):591–597.
27. Tinkelman DG, Price DB, Nordyke RJ, et al. Symptom-based questionnaire for differentiating COPD and asthma. *Respiration*. 2006;73(3):296–305.
28. Lee YS, Baek S, Ko Y, et al. New scoring system for the differentiation of chronic obstructive pulmonary disease and asthma. *Respirology*. 2015;20(4):626–632.
29. Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. *Eur Respir J*. 2005;26(5):948–968.
30. Sin DD, Miravittles M, Mannino DM, et al. What is asthma-COPD overlap syndrome? Towards a consensus definition from a round table discussion. *Eur Respir J*. 2016;48(3):664–673.
31. Plaza V, Alvarez F, Calle M, et al. Consensus on the asthma-COPD overlap syndrome (ACOS) between the Spanish COPD Guidelines (GesEPOC) and the Spanish Guidelines on the Management of Asthma (GEMA) [Consenso sobre el solapamiento de asma y EPOC (ACO) entre la Guía española de la EPOC (GesEPOC) y la Guía Española para el Manejo del Asma (GEMA)]. *Arch Bronconeumol*. 2017;53(8):443–449.
32. Miravittles M. Diagnosis of asthma-COPD overlap: the five commandments. *Eur Respir J*. 2017;49(5).

## Supplementary materials

### Spirometry

This study used spirometry data obtained as part of enrollment in GSK Studies 200699 (NCT02164539; data were recorded using the ERT Masterscope CT) or 201496 (NCT02299375), or obtained as part of routine clinical management within the 6 months prior to the patient providing consent for study inclusion (data were not accessed until the patient had provided written consent for study enrollment). For patients without spirometry data in this period, spirometry was performed at the clinic visit during which the survey was administered.

Data from the following spirometry assessments performed according to accepted American Thoracic Society/European Respiratory Society standards<sup>1</sup> were recorded in all patients: percentage predicted forced expiratory volume in 1 second (FEV<sub>1</sub>), pre- and post-bronchodilator FEV<sub>1</sub> (L), pre- and post-bronchodilator forced vital capacity (FVC) (L); pre- and post-bronchodilator FEV<sub>1</sub>/FVC ratio, and reversibility (mL and %). Reversibility was calculated as follows:

Reversibility in mL

$$= (\text{FEV}_{1 \text{ post-albuterol}} - \text{FEV}_{1 \text{ pre-albuterol}}) \times 1,000$$

Reversibility in %

$$= \frac{\text{FEV}_{1 \text{ post-albuterol}} - \text{FEV}_{1 \text{ pre-albuterol}}}{\text{FEV}_{1 \text{ pre-albuterol}}} \times 100$$

### Statistical considerations

The numbers of patients in each group were based on a preplanned sample size of ~1,000 patients to allow for a broad evaluation of various clinical characteristics associated with asthma and/or chronic obstructive pulmonary disease (COPD).

Data analyses were performed with SAS version 9.3 (SAS Institute, Cary, NC, USA) and R version 3.2.4.<sup>2</sup>

The statistical analyses aimed to identify a primary model (PM) incorporating variables (patient demographics or medical history characteristics based on responses to the patient history questionnaire) that were good predictors of patients' disease categories (asthma only, COPD only, or asthma-COPD overlap [ACO]), and to evaluate the performance of this model in differentiating patients with ACO from those with asthma or COPD as classified by spirometry. Questionnaire items 39, 40, and 41 were intended only for gathering information on the current diagnosis and were therefore not included as explanatory variables for the modeling.

The univariate distribution of each variable was assessed to determine any differences between patient cohorts. Questionnaire items were classified according to the odds ratios (ORs) of the COPD and the asthma cohorts to the ACO cohort, using the first answer listed for each question as the reference, as follows:

- Type 1: ORs of the COPD and asthma cohorts to the ACO cohort both significantly different from 1 in the same direction (both >1 or <1), demonstrating a different response for the ACO cohort to both the asthma and COPD cohorts with the ACO cohort at one end of the spectrum of responses.
- Type 2: ORs both significantly different from 1 in different directions (one <1 and the other >1), demonstrating differences in responses between all cohorts, with the ACO cohort falling between the asthma and COPD cohorts in the spectrum of responses.
- Type 3a: OR significantly different from 1 for comparison with COPD only, demonstrating that the ACO cohort had similar responses to the asthma cohort but different responses from the COPD cohort.
- Type 3c: OR significantly different from 1 for comparison with asthma only, demonstrating that the ACO cohort had similar responses to the COPD cohort but different responses from the asthma cohort.
- Type 4: ORs showed no significant difference between ACO and asthma or COPD, demonstrating similar responses in all cohorts.

To assess the associations between categorical item response variables, correlation coefficient analysis was performed based on Spearman rank-order correlation coefficients. Variables that were moderately or highly correlated (correlation coefficient  $\geq 0.5$ ) were assessed for possible multicollinearity and excluded as appropriate from the multinomial logistic regression analysis. For continuous variables, similar correlation analysis was performed using Pearson's product-moment correlation coefficients.

Multinomial logistic regression analysis was then performed to identify a subset of variables able to predict patients' disease classification, as identified using spirometry, with reasonable accuracy and repeatability. Seven orthogonal partitions of the evaluable population were created using a prespecified random scheme (further details are available in the protocol for this study, which is available online from the GSK Clinical Study Register [number 201703]<sup>3</sup>). For each partition, a split-sample approach was employed for cross-validation, using 50% of data to identify significant predictors and the other 50% for performance evaluation.

Age, sex, body mass index (BMI), smoking status, and exacerbation history were of clinical relevance and included in every analysis model in the variable selection process. Multinomial logistic regression analysis was performed using PROC LOGISTIC in SAS, with ACO as the reference category, using the maximum likelihood method. For each round of modeling, prespecified stepwise variable selection criteria were used at a significance level of 0.10 for a variable to enter the model and 0.15 to remain in the model.

In assessing the performance of the model, the predicted probabilities for each patient in the validation dataset were calculated for the three disease cohorts (asthma only, ACO, or COPD only) based on the fitted model, and the predicted outcome for the patient was determined as the disease cohort with the maximum predicted probability. The disease classifications by spirometry and by model-based prediction were compared.

## Post hoc analyses

To further understand the results from the PM, post hoc analyses were performed to identify variables that differentiate patients in the asthma cohort from all other patients (included in the model for asthma) and patients in the COPD cohort from all other patients (included in the model for COPD). As for the PM, binary logistic regression analysis was performed to identify variables able to discriminate between patients with or without asthma, and with or without COPD. Three versions of the asthma and COPD models were identified: a basic model, a reduced model, and a full model. The basic model for asthma included only age, sex, smoking status, and BMI as explanatory variables, while the basic model for COPD included age, sex, smoking status, and exacerbation history. The reduced models comprised the basic model characteristics plus those found to be significant predictors of disease in at least four out of seven rounds of logistic regression modeling, and the full models

comprised the basic model characteristics plus responses to questionnaire items 1–36, with the exception of item 30. Receiver operating characteristic curves were generated to compare the effectiveness of the three versions of the models. The predicted outcomes using the full models for asthma and COPD were used to classify patients into four groups according to the presence or absence of asthma and COPD (asthma without COPD, COPD without asthma, neither asthma nor COPD, or asthma and COPD). A 100% stacked bar chart was generated to compare these predictions with spirometric disease classification.

Partial least squares discriminant analysis (PLS-DA) was conducted using the R package “ropls”<sup>4</sup> with the input variables of age, sex, smoking status, BMI, exacerbation incidence, and the responses to questionnaire items 1–36, to further examine the degree of separation between the groups identified as having asthma and having COPD, and the groups identified as having asthma, COPD, and ACO, by spirometric classification. Finally, data on percentage reversibility at screening, post-bronchodilator FEV<sub>1</sub>/FVC ratio, and pre-bronchodilator FEV<sub>1</sub> were added to the input variables, and the PLS-DA was performed again. The outcome of this analysis was  $Q^2$ , a cross-validated equivalent of  $R^2$ , which ranges from negative to 1, with 1 representing perfect discrimination.

## References

1. Miller MR, Hankinson J, Brusasco V, et al; ATS/ERS Task Force. Standardisation of spirometry. *Eur Respir J*. 2005;26(2):319–338.
2. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016. Available from: <https://www.R-project.org>. Accessed May 15, 2017.
3. GlaxoSmithKline. The utility of a clinical questionnaire to identify subjects with features of both asthma and COPD. 2016. Available from: [https://www.gsk-clinicalstudyregister.com/search/?study\\_ids=201703](https://www.gsk-clinicalstudyregister.com/search/?study_ids=201703). Accessed June 9, 2016.
4. Thevenot EA, Roux A, Xu Y, Ezan E, Junot C. Analysis of the human adult urinary metabolome variations with age, body mass index and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *J Proteome Res*. 2015;14:3322–3335.

### 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.