Identification of Undiagnosed COPD Patients in a Claims Database Using a Predictive Model

Saverno KR[1], Zhou Y[1], Moretz C[1], Renda A[2], Burslem K[3], Jain G[3], Hernandez G[3], Dhamane A[3]

[1]Comprehensive Health Insights Inc., Louisville, KY; [2]Humana Inc., Louisville, KY; [3]Boehringer Ingelheim Pharmaceuticals Inc. (BIPI), Ridgefield, CT

Background

- Chronic obstructive pulmonary disease (COPD) is a leading cause of morbidity and mortality and results in substantial social and economic burden.[1-4]
- In the United States (US), it is estimated that of the 26.8 million people with chronic obstructive pulmonary disease (COPD), 45% remain undiagnosed.[5] The estimated direct costs of COPD are $29.5 billion and the indirect costs are $20.4 billion, with exacerbations accounting for the greatest proportion of total COPD burden.[5]
- Identification of patients with undiagnosed COPD is essential for implementing interventions, such as smoking cessation and drug therapies, which are aimed at improving outcomes and preventing disease progression.[2,6]
- A previously published predictive model to identify patients at risk for having undiagnosed COPD exhibited a low positive predictive value (24.9%) and limited generalizability (Health Maintenance Organization [HMO] serving New Mexico).[7]
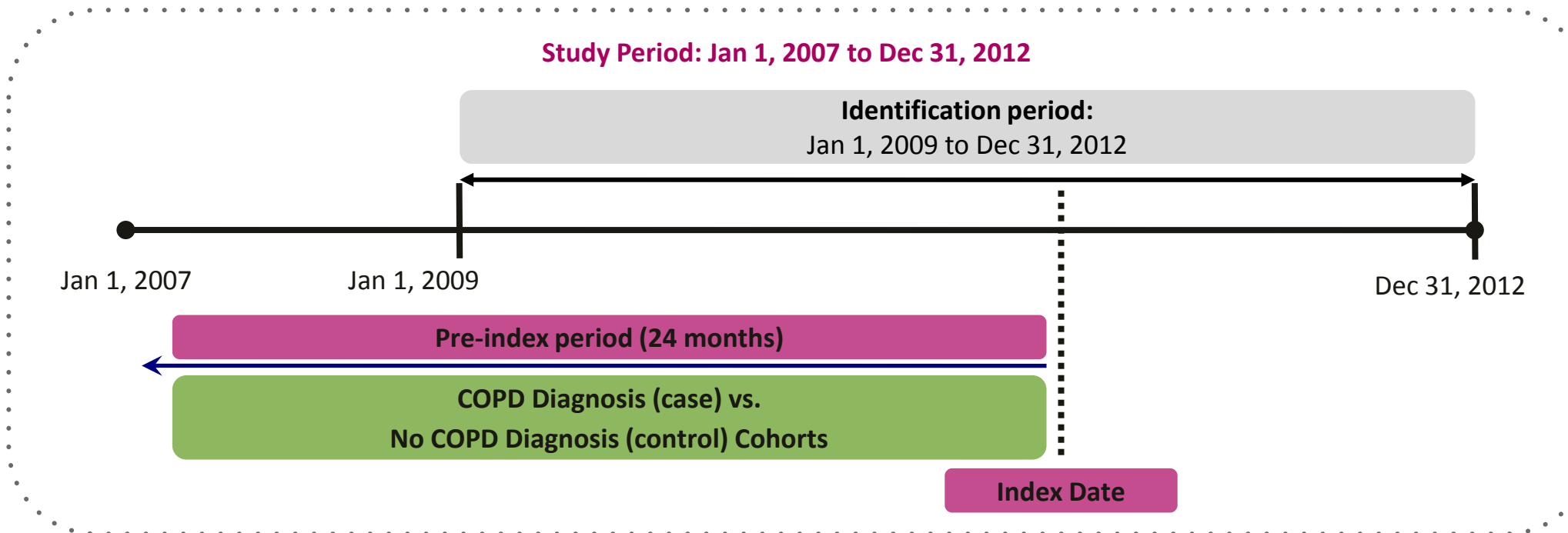
Objective

To develop and validate a predictive model for identifying patients likely to have undiagnosed COPD, using a large retrospective claims database.

Methods

- **Study Design:** Retrospective cohort study.
- **Figure 1. Study Design**

Study Period: Jan 1, 2007 to Dec 31, 2012

Identification period: Jan 1, 2009 to Dec 31, 2012

Jan 1, 2007 | Jan 1, 2009 | Dec 31, 2012

Pre-index period (24 months)

COPD Diagnosis (case) vs. No COPD Diagnosis (control) Cohorts

Index Date

- **Data Source:** The Humana administrative claims database, which contains integrated medical, pharmacy, behavioral, and lab-related claims, and includes patients enrolled in commercial, Medicare advantage, and prescription drug plans.
- **Patient selection:**
  - Two cohorts of individuals with (cases) and without (controls) a COPD diagnosis were identified between January 1, 2009 and December 31, 2012.
    - COPD diagnosis was defined as ≥2 medical claims (separated by 90 days) with a COPD diagnosis code [ICD-9-CM (International Classification of Diseases, 9th revision, Clinical Modification) codes 491.xx (chronic bronchitis), 492.xx (emphysema), or 496.xx (COPD, unspecified)] in the primary or secondary position during the identification period. *The date of the first medical claim with a COPD diagnosis code was designated as the index date.*
  - Included patients were aged 40-89, enrolled in a Medicare Advantage with Prescription Drug coverage (MAPD) or commercial health plan, with a minimum of two years of continuous enrollment.
  - Patients with a diagnosis code for cystic fibrosis (277.0x), pulmonary tuberculosis (011.x), or malignant neoplasms (140.xx-172.xx, 174.xx-209.3x, or 209.7x) anytime during the study period were excluded.
  - Stratified random sampling based on line of business (LOB) and year of index date was used to select one control for each case, to ensure that the two cohorts had similar distributions of these variables.
- **Predictive Model Development:**
  Model Development
  - A total of 61 demographic and clinical characteristics were assessed during the pre-index period (full list available from authors upon request).
    - Demographic characteristics included age, gender, geographic region, and LOB.
    - Clinical characteristics, including comorbidities and healthcare resource utilization, were identified through a review of the published literature[7-13] and from the opinions of Humana's and BIPI's clinical experts

GCHHWAGEN

- **Predictive Model Development (continued):**
  Model Development
  - The two cohorts (cases and controls) were compared to identify characteristics which could be predictive of COPD diagnosis.
    - Analysis of Variance (ANOVA) was used for continuous variables and Chi-square for categorical variables
  - Three parallel modeling approaches were pursued for variable selection: neural networking, logistic regression (backward and stepwise), and decision tree.
    - Variable selection was performed based on model-specific selection criteria (average squared error, profit/loss function).
  - The models were trained, validated, and tested on randomly portioned subsets of the sample (40%, 30%, and 30% respectively).
  Model Selection
  - The final model was selected based on the area under the curve (AUC) index of the Receiver Operating Characteristic (ROC) curve of the test subset.
    - AUC index was used as the model evaluation metric because of its ability to indicate how well the model discriminates between cases and controls.

Table 1. Patient Selection

Criteria	Cases Remaining (Patients with COPD)		Controls Remaining (Patients without COPD)	
	n	%	n	%
INCLUSION				
Medical claim for COPD during the identification period (Jan 1, 2009 to December 31, 2012)	700,664	100.0%		
More than 24 months of continuous enrollment with both medical and pharmacy benefits during the identification period (Jan 1, 2009 to December 31, 2012).			3,753,786	100.0%
EXCLUSION				
The second COPD claim does not occur within 90 days of the first COPD claim	353,738	50.5%		
Diagnosis of COPD during the 24-month pre-index period	267,143	38.1%	3,139,883	83.6%
Diagnosis of cystic fibrosis, pulmonary tuberculosis, or malignant neoplasms during the study period (Jan 1, 2007 to December 31, 2012)	200,845	28.7%	2,829,467	75.4%
Not aged 40-89 years on the index date	192,665	27.5%	1,802,705	48.0%
No continuous enrollment with pharmacy and medical benefits for a 24-month pre-index period	50,880	7.3%		
STUDY SAMPLE AVAILABLE FOR MATCH	**50,880**	**100%**	**1,802,705**	**100%**

Results

Table 2. Baseline Demographics

Characteristic	Cases (Patients with COPD)	Controls (Patients without COPD)	p-value
	n = 50,880	n = 50,880	
Age, years			
Mean (SD)	71.36 (10.1)	68.27 (9.9)	<.0001
Gender - n (%)			
Male	23,585 (46.4%)	21,996 (43.2%)	<.0001
Female	27,295 (53.6%)	28,884 (56.8%)	
Race - n (%)			
White	39,836 (78.3%)	38,672 (76.0%)	
Black	4,823 (9.5%)	5,288 (10.4%)	
Hispanic	966 (1.9%)	857 (1.7%)	<.0001
Other	1,160 (2.3%)	1,931 (3.8%)	
Unknown	4,095 (8.0%)	4,132 (8.1%)	
Geographic Region - n (%)			
Northeast	893 (1.8%)	1,219 (2.4%)	
Midwest	12,584 (24.7%)	13,915 (27.3%)	<.0001
South	33,262 (65.4%)	29,116 (57.2%)	
West	4,141 (8.1%)	6,630 (13.0%)	
Plan Characteristics - n (%)			
LIS Status	8,406 (16.5%)	5,473 (10.8%)	<.0001
Dual Eligibility	5,684 (11.2%)	3,703 (7.3%)	<.0001
Payer Type - n (%)			
Commercial	4,056 (8.0%)	4,056 (8.0%)	<.0001
Medicare	46,824 (92.0%)	4,6824 (92.0%)	
Deyo Charlson Comorbidity Index			
Mean (SD)	1.73 (2.0)	0.87 (1.4)	<.0001

SD, standard deviation

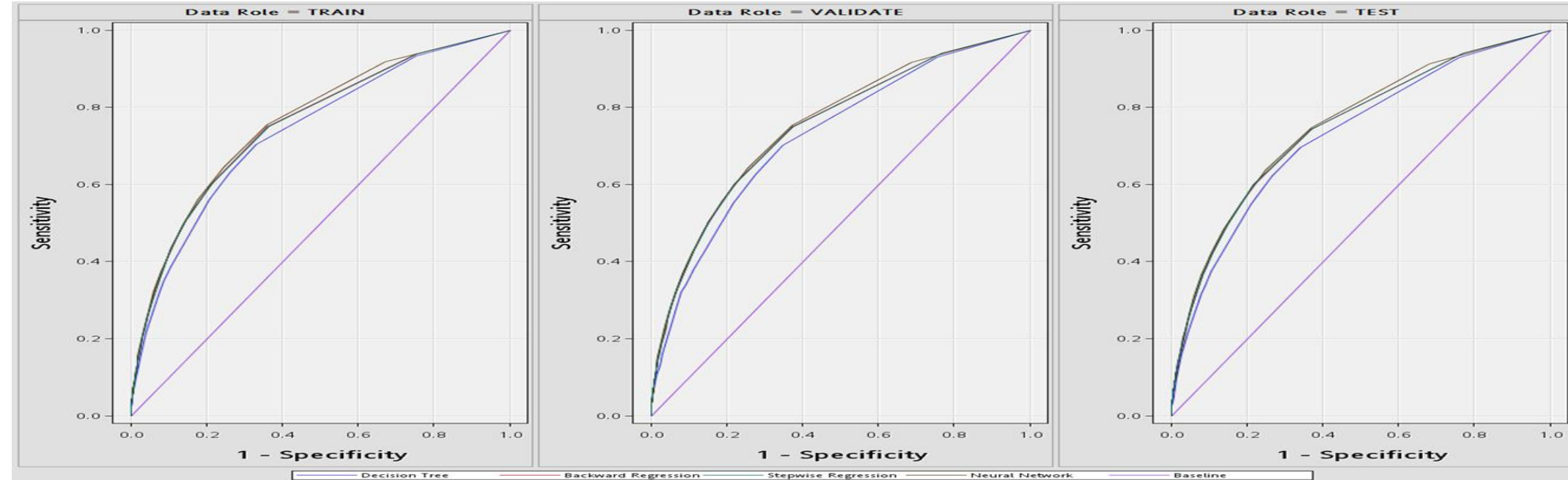### Figure 2. Model Comparison using Receiver Operator Characteristic Curves

Table 3. Application of the Stepwise Logistic Regression Model to the Test Subset

		Diagnosis According to Study Criteria		Total
		COPD	No COPD	
Model Prediction	COPD	9,151 True Positives	3,378 False Positives	12,529
	No COPD	6,113 False Negatives	11,885 True Negatives	17,998
	Total	15,264	15,263	30,527

Positive Predictive Value (PPV) = (9,151/12,529)*100 = 73%
Negative Predictive Value (NPV) = (11,885/ 17,998)*100 = 66%
Sensitivity = (9,151/15,264)*100 = 60%
Specificity = (11,885/15,263)*100 = 78%
Area Under the Curve = 0.754

Figure 3. Predictive Model (Stepwise Logistic Regression): Odds Ratio (95% CI)

Variable	Odds Ratio (95% CI)	
Anticholinergic bronchodilators	3.34 (2.35 - 4.73)	
Tobacco cessation counseling	2.88 (2.67 - 3.09)	
Anticholinergic beta-agonist combination agents	2.68 (2.12 - 3.37)	
Smoking cessation medications	2.32 (1.96 - 2.73)	
Region=South vs. West	2.02 (1.92 - 2.13)	
LABA/ICS combination	1.80 (1.65 - 1.97)	
Oxygen	1.72 (1.53 - 1.94)	
Region=Midwest vs. West	1.64 (1.57 - 1.70)	
Short acting beta2 agonist	1.62 (1.54 - 1.70)	
Heart failure	1.60 (1.53 - 1.66)	
Respiratory medications	1.53 (1.45 - 1.60)	
Asthma	1.47 (1.42 - 1.52)	
Aortic aneurysm	1.39 (1.33 - 1.46)	
Pneumonia or influenza	1.39 (1.34 - 1.43)	
Asphyxia	1.38 (1.30 - 1.46)	
Atherosclerosis	1.35 (1.32 - 1.38)	
Bronchitis (not chronic)	1.31 (1.12 - 1.34)	
Respiratory symptoms	1.31 (1.29 - 1.33)	
Arterial circulatory disease	1.27 (1.25 - 1.29)	
Ischemic heart disease	1.22 (1.20 - 1.23)	
Antidepressants	1.18 (1.17 - 1.19)	
Depression	1.16 (1.15 - 1.18)	
Deyo Charlson Comorbidity Index	1.13 (1.12 - 1.13)	
RxRisk-V	1.06 (1.06 - 1.06)	
Age	1.04 (1.04 - 1.04)	
Number of total oral corticosteroid prescriptions	1.03 (1.03 - 1.03)	
Cardiovascular medications	0.86 (0.85 - 0.87)	
Diabetes	0.85 (0.84 - 0.86)	
Hypertension	0.83 (0.82 - 0.84)	
Vaccination/medication to treat influenza	0.82 (0.81 - 0.83)	
Hospitalization (all-cause)	0.80 (0.79 - 0.81)	
Cardio-pulmonary Exercise Test	0.79 (0.78 - 0.81)	
Pneumococcal vaccination	0.76 (0.74 - 0.78)	
Gender=Female vs. Male	0.73 (0.72 - 0.74)	
Oral corticosteroid	0.69 (0.66 - 0.72)	
Intercept	0.02 (0.01 - 0.03)	

0 1 2 3 4 5

CI, confidence interval; LABA, long-acting beta-agonist; ICS, inhaled corticosteroid

Results Summary

- A total of 50,880 cases and 1,802,705 controls were identified (Table 1). Stratified random sampling provided 50,880 matched pairs.
- The models generated by neural networking and stepwise logistic regression (SLR) performed similarly in terms of model diagnostics (AUC), (Figure 2).
  - Neural networking model AUC = 0.757 ; SLR AUC = 0.754.
  - The SLR approach was used to generate the final predictive model, as model performance was similar and SLR is more widely applied than neural networks.
- The final SLR predictive model had sensitivity of 60%, specificity of 78%, PPV of 73%, and NPV of 66% (Table 3). It included 34 variables that were statistically significantly associated with COPD diagnosis (Figure 3).
- Factors with the strongest influence on the predictive model include: anticholinergic bronchodilators, tobacco cessation counseling, anticholinergic/beta-agonist combination, smoking cessation medications and geographic region (Figure 3).

Discussion

- The current model compares favorably to a previously published model.
  - The earlier Mapel et al[7] model had a sensitivity of 60.5%, specificity of 82.1%, a PPV of 24.9% and a NPV of 95.5%.
  - Although the Mapel et al[7] model has higher NPV, PPV is higher in the current model, providing greater efficiency for screening those with a positive result.
  - The current model may have greater generalizability as it was conducted in a large national health plan.
- The differences in model performance may reflect differences in:
  - Clinical characteristics assessed (e.g. use of a composite "Respiratory Rx" variable in Mapel et al[7] compared to variables differentiating respiratory drug classes in the current study) and study populations, including differences in COPD prevalence.

Conclusion

The predictive model may provide a valuable method to target those likely to have COPD, enabling timely diagnosis and appropriate treatment and preventing future events that require valuable resources.

Limitations

- This study is subject to limitations including coding errors of omission and commission, incomplete claims, unreliable clinical coding, and unobservable factors that may also influence the outcomes.
- COPD diagnosis was assessed using administrative claims data and clinical accuracy of diagnosis was not confirmed based on chart review or spirometry results.

References

1. World Health Organization. Chronic Obstructive Pulmonary Disease (COPD). 2013.
2. Global Initiative for Chronic Obstructive Lung Disease, I. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. 2013.
3. Ford ES, et al., Chronic Obstructive Pulmonary Disease Surveillance-US, 1999-2011. Chest, 2013.
4. United States Department of Health and Human Services. National Institutes of Health. Chronic Obstructive Pulmonary Disease. 2013.
5. National Heart, Lung, and Blood Institute. Morbidity and mortality chartbook on cardiovascular, lung and blood disease. Bethesda, Maryland: US Department of Health and Human Services, Public Health Service, National Institutes of Health. 2009.
6. Kohansal R et al. Am J Respir Crit Care Med. 2009;180(1):3-10.
7. Mapel DW et al. J Manag Care Pharm. 2006;12(6):457-65.
8. Mapel DW et al. BMC Health Serv Res. 2011;11:43.
9. Mannino DM et al. Eur Respir J. 2008;32(4):962-9.
10. Bartlett JG et al. Management of infection in acute exacerbations of COPD. 2013.
11. Mosenifar Z et al. Chronic Obstructive Pulmonary Disease. 2013 May 28, 2013.
12. Sharafkhaneh A et al. Int J Chron Obstruct Pulmon Dis. 2010;5:125-32.
13. Watson L et al. Respir Med. 2006;100(4):746-53.