Prediction of Acute Respiratory Disease in Current and Former Smokers With and Without COPD

Russell P. Bowler, MD, PhD; Victor Kim, MD; Elizabeth Regan, MD; André A. A. Williams, PhD; Stephanie A. Santorico, PhD; Barry J. Make, MD, FCCP; David A. Lynch, MD; John E. Hokanson, PhD; George R. Washko, MD; Peter Bercz, MD; Xavier Soler, MD; Nathaniel Marchetti, MD, FCCP; Gerard J. Criner, MD, FCCP; Joe Ramsdell, MD, FCCP; MeiLan K. Han, MD; Dawn Demeo, MD; Antonio Anzueto, MD; Alejandro Comellas, MD; James D. Crapo, MD, FCCP; Mark Dransfield, MD; J. Michael Wells, MD; Craig P. Hersh, MD; Neil MacIntyre, MD, FCCP; Fernando Martinez, MD, FCCP; Hrudaya P. Nath, MD, FCCP; Dennis Niewoehner, MD, FCCP; Frank Sciurba, MD, FCCP; Amir Sharafkhaneh, MD, FCCP; Edwin K. Silverman, MD, PhD; Edwin J. R. van Beek, MD; Carla Wilson; Christine Wendt, MD; and Robert A. Wise, MD, FCCP; for the COPDGene investigators

e-Appendix 1.

Supplemental Methods

Random Forests Methodology

Random survival forests were used to analyze the data with time to severe episodes as the outcome of interest. One hundred and forty two variables were considered in the analysis. Each forest contained 500 trees. The log rank splitting rule was used to grow the trees. A Bootstrap resample of the data was used for each tree. Each resample contained approximately two thirds of the data. The other 'left out 'or out of bag (OOB) samples were used to calculate the OOB error rate. The number of variables assessed, at each branch of a tree, was a random selection of p1/2 where p is the total number of variables. The splitting continues until a terminal node reaches a minimum number of 3 unique times to severe exacerbation. Missing data was not imputed. The analysis was performed separately for subjects with GOLD status 2 through 4; 0 and 1; and unclassified subjects. For each analysis the variable importance is reported.

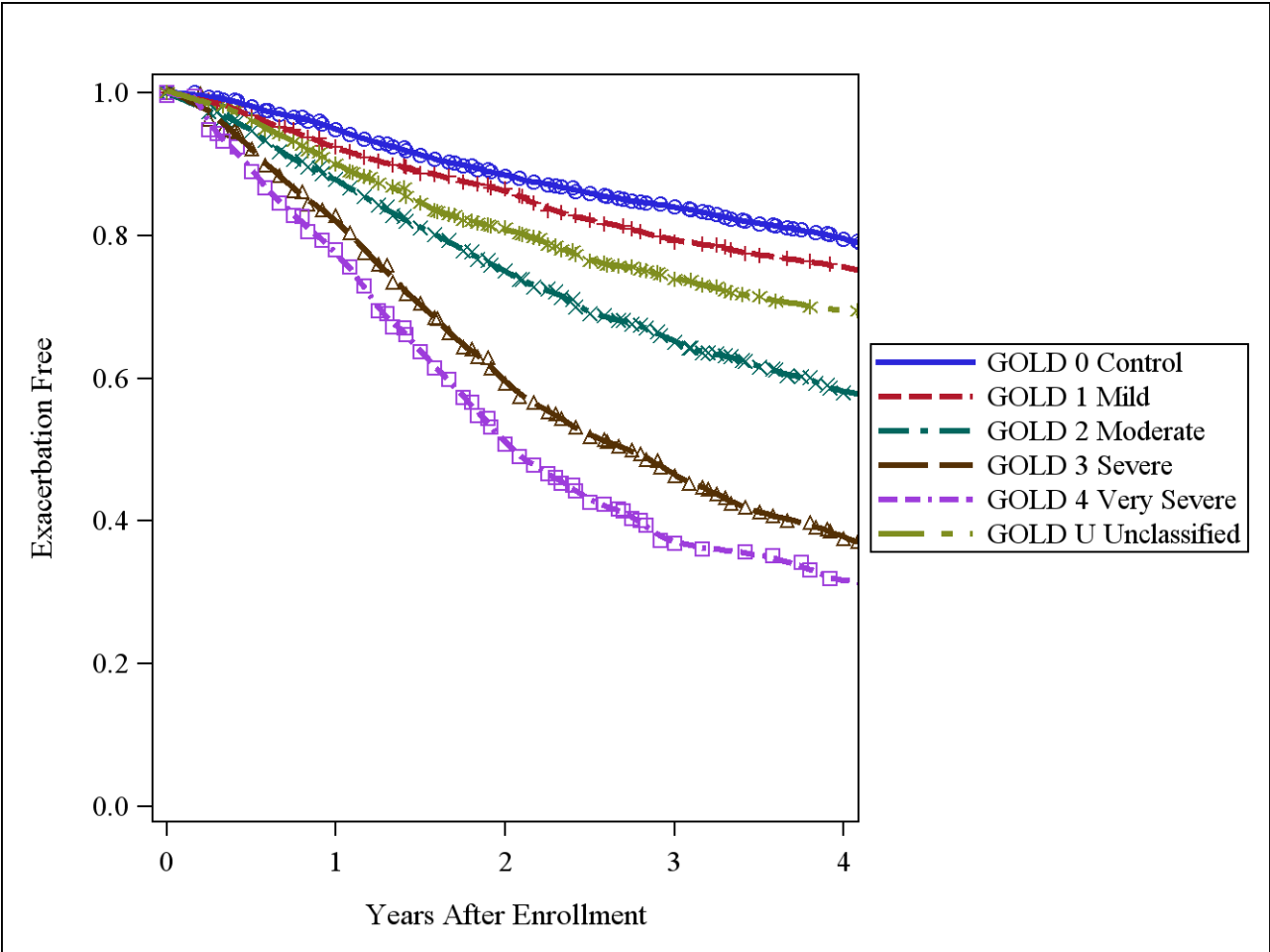Acute episodes of respiratory disease Score

Cox proportional hazards modeling multiple regression was used to identify parameter estimates ($\hat{\beta}_i$) for the three major variables (x_i) that were independently predictive of acute episodes of respiratory disease in the LFU period (SGRQ, FEV1%, prior acute episodes of respiratory disease). The entire dataset was used to calculate parameter estimates. The baseline survival function s(t) and the parameter estimates of each independent variable were used to estimate the probability that a subject will have an

acute episodes of respiratory disease over a given time t (in years) for a subject j is $P_j(t) = \dfrac{s(t)e^{\sum_{n=1}^{N}(\hat{\beta}i * \mathrm{x}ij)}}{s(t)e^{\sum_{n=1}^{N}(\hat{\beta}i * \mathrm{x}ij)} + 1}$.

A preliminary score ($Sp = \sum_{n=1}^{N}(\hat{\beta}i * \mathrm{x}ij)$) was transformed into a more intuitive exacerbation score by linear mapping onto a scale of 0-20. Each component score was rounded to a whole number.

Additional Statistical Methods

Several strategies were used to determine important variables that might be associated with acute episodes of respiratory disease including random forest analysis (**e-Table 1**), univariate regression modeling (**e-Table 2**), and risk factors identified in the literature. A visual inspection of the correlation matrix for all variables was used to determine collinearity among variables. When variables were highly collinear, we used the variable with the strongest univariate association. For multivariate modeling, the following covariates were used: demographics and physical characteristics (age, race, gender, BMI,

height); medical history (current versus former smoker, pack years, history of gastroesophageal reflux (GERD), years of smoke exposure at work, history of working at a dusty job, history of COPD in a parent, use of oxygen, history of congestive heart failure, history of blood clots); chronic bronchitis, exacerbations in the previous year, whether the 6-minute walk test was limited by dyspnea; physiology (FEV_1% predicted and FEV_1/FVC ratio post bronchodilator, bronchodilator reversibility as defined by the ATS and European Respiratory Society[24], resting oxygen saturation, 6-minute walk distance); questionnaires (St. George's Respiratory Questionnaire (SGRQ) and Modified Medical Research Council (MMRC) dyspnea score); and imaging (emphysema, gas trapping, PA, Ao, and PA/Ao).

e-Figure 1: Fraction of subjects remaining episode free by GOLD stage

e-Table 1: Random forest analysis of acute episodes of respiratory disease frequency

Variable Name	Variable importance	Variable Name	Variable importance	Variable Name	Variable importance
Exacerbation in prior 12 month	1	bronchitis Tracheobronchomalacia on CT	0.011	bronchiectasis on CT History of stomach ulcers	-0.002
History Bronchitis Attack	0.22	Peripheral vascular disease	0.008	Sleep apnea	-0.002
Height (cm)	0.129	Mother with emphysema	0.008	Back pain	-0.003
History of Pneumonia	0.106	Other high concerns on Ct	0.007	High blood pressure	-0.003
SGRQ (symptom)	0.094	Total lung capacity on CT	0.007	Father smoked cigarettes	-0.003
Self Reported Health Status (poor-excellent)	0.065	BODE	0.006	Gastroesophageal reflux	-0.004
History of chronic bronchitis	0.06	Father with COPD	0.006	Functional residual capacity on CT	-0.005
Age at enrollment	0.059	Concerns for ILD on CT	0.005	CABG	-0.006
History of asthma	0.053	Mother smoked cigarettes	0.005	Blood pressure (systolic)	-0.006
Mosiac attenuation on CT	0.046	Race	0.005	Usually coughs up phlegm	-0.007
History of breast cancer	0.041	Honeycombing on CT	0.004	Concern for cancer on CT	-0.007
Distance walked in 6 min.	0.041	History of colon cancer	0.004	Mother with lung cancer	-0.007
Pi15 (square root of WA)	0.036	Compression fractures in back	0.004	Atelectasis on CT	-0.007
History of COPD	0.036	Forced vital capacity	0.003	Heart attack	-0.008
Use oxygen	0.035	Wheezing or whistling in chest	0.003	Change in FEV_1 post BD	-0.008
Bronchial wall thickening on CT	0.033	Angina	0.003	High cholesterol	-0.008
Gender	0.029	%change in FEV_1 post BD	0.003	DustyJobEver	-0.009
Father with Emphysema	0.028	Mother with asthma	0.002	TIA	-0.009
History of pneumothorax	0.019	Gas trapping noted on CT	0.001	Cancer of prostate	-0.009
Diagnosis of emphysema	0.018	History of blood clots	0.001	ChestIllOther	-0.01
History of osteoporosis	0.017	History of bladder cancer	0.001	Cysts on CT	-0.01
Joint pain/stiffness limiting walking	0.016	History of congestive heart failure	0.001	Macular degeneration	-0.01
History allergic conjunctivitis	0.016	Coronary artery disease	0.001	Hay fever	-0.01
No high concerns on CT	0.016	History of hip fracture	0	ATS pack years	-0.01
History of stroke	0.015	Osteoarthitis	0	ATS chronic bronchitis	-0.012
History of diabetes	0.015	Ground glass on CT	0	Consolidation on CT	-0.013
Bronchiectasis on CT	0.015	Noncalcificed nodule CT	0	BD responsive (ATS)	-0.014
Blood Pressure (diastolic)	0.015	History of rhinitis	-0.001	Peak expiratory flow	-0.014
Usually has cough	0.015	Pi10 (square root of WA)	-0.001	ATS/ERS classification	-0.015
Mother with COPD	0.013	Smokes cigars regularly	-0.001	Resting SaO2	-0.016
Father with asthma	0.012	Smokes pipe regularly	-0.001	Reticular abnormality CT	-0.016
Father with chronic Bronchitis	0.012	Rheumatoid arthitis	-0.002	Oxygen used on 6 m walk	-0.018
Mother with chronic	0.011	Concerns of	-0.002	Father with lung cancer	-0.018
				SGRQ (active)	-0.02
				Pre BD PEF	-0.021
				Pre BD FVC	-0.03
				Angioplasty	-0.03
				Dyspnea limiting walking	-0.03

Variable Name	Variable importance	Variable Name	Variable importance	Variable Name	Variable importance
WA% (segmental)	-0.031	Emphysema (upper 1/3)	-0.064	Emphysema (upper 1/3)	-0.086
Emphysema (%)	-0.037	BMI	-0.071	FEV_1	-0.092
MMRC dyspnea score	-0.038	SGRQ (impact)	-0.071	Pre BD FEF 25/75	-0.101
Change in FVC post BD	-0.048	FEF 25/75 post BD	-0.078	FEV_1 pre BD	-0.104
%change FVC post BD	-0.048	FEV_1 /FVC	-0.08	Pre BD FEV_1 /FVC	-0.134
Weight	-0.054	Upper/lower emphsyema	-0.081		
Gas trapping on CT	-0.06	SGRQ (total)	-0.082		

BD: bronchodilator; SGRQ:(St. George's Respiratory Questionnaire); PEF: peak expiratory flow; FVC: forced vital capacity; FEV_1: forced expiratory volume at 1 second; WA: wall area on CT;

e-Table 2: Univariable associations with acute episodes of respiratory disease frequency 12 months prior to enrollment

	Negative Binomial			Zero Model		
Risk Factor	**Estimate**	**SE**	**P**	**Estimate**	**SE**	**P**
Age at enrollment	-0.002063	0.004676	0.6590	-0.142376	0.048014	0.0030
Gender (female)	0.007452	0.079085	0.9249	1.098272	0.548693	0.0453
Race (NHW)	-0.049916	0.095704	0.6020	-1.046209	0.409298	0.0106
GERD	-0.212289	0.079241	0.0074	2.206574	2.007017	0.2716
CAD	-0.009484	0.104702	0.9278	1.901872	2.206111	0.3886
CHF	-0.388110	0.137490	0.0048	18.938039	0.116097	<.0001
Current smoker	0.230205	0.081148	0.0046	-0.640178	0.292653	0.0287
ATS pack years	0.001480	0.001403	0.2917	-0.045439	0.012979	0.0005
6 min walk with O_2	-0.438017	0.073059	<.0001	22.380420	0.046925	<.0001
Chronic bronchitis	-0.407711	0.080057	<.0001	2.692825	2.786145	0.3338
History of asthma	-0.296329	0.122025	0.0152	2.683252	2.124752	0.2066
Parent with COPD	-0.178017	0.093734	0.0575	3.917705	15.628784	0.8021
Supplemental O_2	-0.438017	0.073059	<.0001	22.380420	0.046925	<.0001
BP (systolic)	-0.004154	0.002309	0.0720	-0.000957	0.005971	0.8726
BP (diastolic)	-0.005178	0.003410	0.1289	0.022456	0.011093	0.0429
Heart rate	0.009235	0.002820	0.0011	-0.040516	0.009849	<.0001
BMI	0.004488	0.004928	0.3625	-0.013050	0.011623	0.2615
6 minute walk	-0.000465	0.000092721	<.0001	0.002617	0.000264	<.0001
FEV_1	-0.010408	0.001337	<.0001	0.044956	0.002899	<.0001
FEV_1/FVC	-1.188362	0.187060	<.0001	6.990424	0.501006	<.0001
BDR	-0.013018	0.083014	0.8754	17.014059	0.082823	<.0001
BODE	0.151846	0.017379	<.0001	-1.045358	0.095608	<.0001
MMRC	0.260451	0.028110	<.0001	-0.988874	0.118842	<.0001
SGRQ (total)	0.019799	0.001724	<.0001	-0.078528	0.006519	<.0001
Emphysema (%)	0.008658	0.002603	0.0009	-0.131201	0.014576	<.0001

	Negative Binomial			Zero Model		
Risk Factor	Estimate	SE	P	Estimate	SE	P
gas trapping (%)	0.006801	0.001620	<.0001	-0.048072	0.004621	<.0001
TLC (%)	0.007471	0.001971	0.0002	-0.020808	0.003737	<.0001
Pi10	0.875251	0.265725	0.0010	-5.684710	0.676708	<.0001
WA% (segmental)	0.035231	0.011877	0.0030	-0.272665	0.033964	<.0001
WA% (subsegmental)	0.041324	0.021976	0.0601	-0.424565	0.062523	<.0001
PA (mm)	0.013531	0.008968	0.1313	-0.028886	0.022015	0.1895
Ao (mm)	-0.003465	0.014415	0.8100	0.016866	0.044500	0.7047
PA/Ao	0.528598	0.322090	0.1008	-1.270474	0.811970	0.1177

SE: standard error; NHW: non-Hispanic White; GERD: gastroesophageal reflux; CAD: coronary artery disease; CHF: congestive heart failure; BP: blood pressure; BMI: body mass index FEV_1: forced expiratory volume at 1 second;; FVC: forced vital capacity; FEV1: BDR: bronchodilator responsiveness; MMRC: Modified Medical Research Council dyspnea score; SGRQ:(St. George's Respiratory Questionnaire); TLC: total lung capacity; Pi10 (square root of wall area percent for 10 μm airway; WA% wall area %; PA: pulmonary artery; Ao: aorta

e-Table 3: Univariable associations with hospitalizations for acute episodes of respiratory disease in 12 months prior to enrollment

RiskFactor	Negative Binomial			Zero Model		
	Estimate	StdErr	Probt	Inf_Est	Inf_Error	Inf_P
Age at enrollment	0.013681	0.006140	0.0259	0.771384	0.530673	0.1461
Gender (femal)	0.068063	0.184686	0.7125	17.574720	0.885842	<.0001
Race (NHW)	-0.231853	0.108365	0.0324	1.879511	.	.
GERD	-0.098212	0.191598	0.6082	18.177062	0.333609	<.0001
CAD	-0.248083	0.220633	0.2608	18.510602	0.312956	<.0001
CHF	-0.797402	0.246703	0.0012	21.241856	0.172784	<.0001
Current smoker	0.154249	0.192241	0.4223	-16.481721	.	.
ATS pack years	0.004228	0.002617	0.1062	-0.032141	0.012903	0.0127
6 min walk with O_2	-0.238233	0.166605	0.1527	22.885925	0.089201	<.0001
Chronic bronchitis	-0.491576	0.190875	0.0100	19.370001	0.266717	<.0001
History of asthma	-0.506666	0.304024	0.0956	0.753054	0.634052	0.2350
Parent with COPD	0.105878	0.211768	0.6171	16.803944	0.182431	<.0001
Supplemental O_2	-0.238233	0.166605	0.1527	22.885925	0.089201	<.0001
BP (systolic)	-0.000270	0.003739	0.9424	0.129826	0.086315	0.1326
BP (diastiolic)	-0.006130	0.007529	0.4155	0.044913	0.078031	0.5649
Heart rate	0.012699	0.006216	0.0410	-0.050437	0.010418	<.0001
BMI	0.009574	.	.	-0.010093	171.659494	1.0000
6 min walk (ft)	-0.000896	0.000235	0.0001	0.003307	0.000431	<.0001
FEV_1	-0.008395	0.003269	0.0102	0.052157	0.004948	<.0001
FEV1_FVC_utah	-0.760115	0.444981	0.0876	7.069481	0.781050	<.0001
BDR 0	-0.223785	0.190728	0.2407	18.664741	0.235369	<.0001
BODE	0.238619	0.046558	<.0001	-1.215301	0.196447	<.0001
MMRC	0.544481	0.091881	<.0001	-1.044378	0.297948	0.0005
SGRQ (total)	0.035160	0.004855	<.0001	-0.111606	0.024041	<.0001
%emphysea	0.008499	0.005788	0.1420	-0.107637	0.019454	<.0001

	Negative Binomial			Zero Model		
RiskFactor	**Estimate**	**StdErr**	**Probt**	**Inf_Est**	**Inf_Error**	**Inf_P**
%gas trapping	0.003737	0.003731	0.3165	-0.048691	0.007176	<.0001
TLC (%)	0.009861	0.004691	0.0356	-0.013681	0.008553	0.1097
Pi10	1.434824	0.580837	0.0135	-6.685049	0.992454	<.0001
WA% (segmental)	0.072145	0.030026	0.0163	-0.316360	0.060134	<.0001
WA% (subsegmental)	0.122448	0.053545	0.0222	-0.915138	0.256028	0.0004
PA (mm)	0.057920	0.022735	0.0108	-0.062984	0.053902	0.2426
Ao (mm)	0.017803	0.022105	0.4206	0.370665	0.228313	0.1045
PA/Ao	1.800229	0.787941	0.0223	-2.593733	1.998740	0.1944

SE: standard error; NHW: non-Hispanic White; GERD: gastroesophageal reflux; CAD: coronary artery disease; CHF: congestive heart failure; BP: blood pressure; BMI: body mass index FEV_1: forced expiratory volume at 1 second;; FVC: forced vital capacity; FEV1: BDR: bronchodilator responsiveness; MMRC: Modified Medical Research Council dyspnea score; SGRQ:(St. George's Respiratory Questionnaire); TLC: total lung capacity; Pi10 (square root of wall area percent for 10 μm airway; WA% wall area %; PA: pulmonary artery; Ao: aorta

e-Table 4: Annual rate of hospitalization for acute episodes of respiratory disease (actual rates)

Score	Control	GOLD U	COPD GOLD 1,2,3,4
0-2	0.01 ± 0.10 (1020)		0 ± 0 (100)
2-4	0.04 ± 0.27 (1489)	0.03 ± 0.19 (306)	0.02 ± 0.11 (619)
4-6	0.09 ± 0.47 (523)	0.06 ± 0.24 (256)	0.07 ± 0.28 (637)
6-8	0.12 ± 0.57 (268)	0.08 ± 0.33 (192)	0.12 ± 0.39 (701)
8-10	0.48 ± 1.11 (102)	0.12 ± 0.41 (116)	0.21 ± 0.57 (678)
10-12	0.34 ± 0.74 (26)	0.40 ± 0.71 (81)	0.31 ± 0.82 (556)
12-14		0.55 ± 1.17 (35)	0.56 ± 1.08 (292)
>14			0.80 ± 1.34 (220)

means ± standard deviation (#subjects per group; min. 25 subjects per group)