Longitudinal Modeling of Lung Function Trajectories in Smokers with and without Chronic Obstructive Pulmonary Disease

James C. Ross^{1*}, Peter J. Castaldi^{2,3*}, Michael H. Cho^{2,4}, Craig P. Hersh^{2,4}, Farbod N. Rahaghi⁴, Gonzalo V. Sánchez-Ferrero¹, Margaret M. Parker², Augusto A. Litonjua², David Sparrow^{5,6}, Jennifer G. Dy⁷, Edwin K. Silverman^{2,4}, George R. Washko^{4‡}, and Raúl San José Estépar^{1‡}

¹Department of Radiology, ²Channing Division of Network Medicine, ³Division of General Medicine, and ⁴Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston Massachusetts; ⁵VA Normative Aging Study, Veterans Affairs Boston Healthcare System, Boston, Massachusetts; ⁶Department of Medicine, Boston University School of Medicine, Boston, Massachusetts; and ⁷Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts

ORCID ID: 0000-0002-2338-3644 (J.C.R.).

Abstract

Rationale: The relationship between longitudinal lung function trajectories, chest computed tomography (CT) imaging, and genetic predisposition to chronic obstructive pulmonary disease (COPD) has not been explored.

Objectives: 1) To model trajectories using a data-driven approach applied to longitudinal data spanning adulthood in the Normative Aging Study (NAS), and 2) to apply these models to demographically similar subjects in the COPDGene (Genetic Epidemiology of COPD) Study with detailed phenotypic characterization including chest CT.

Methods: We modeled lung function trajectories in 1,060 subjects in NAS with a median follow-up time of 29 years. We assigned 3,546 non-Hispanic white males in COPDGene to these trajectories for further analysis. We assessed phenotypic and genetic differences between trajectories and across age strata.

Measurements and Main Results: We identified four trajectories in NAS with differing levels of maximum lung function and rate of decline. In COPDGene, 617 subjects (17%) were assigned to the lowest trajectory and had the greatest radiologic burden of disease ($P < 0.01$); 1,283 subjects (36%) were assigned to a low trajectory with evidence of airway disease preceding emphysema on CT; 1,411 subjects (40%) and 237 subjects (7%) were assigned to the remaining two trajectories and tended to have preserved lung function and negligible emphysema. The genetic contribution to these trajectories was as high as 83% ($P = 0.02$), and membership in lower lung function trajectories was associated with greater parental histories of COPD, decreased exercise capacity, greater dyspnea, and more frequent COPD exacerbations.

Conclusions: Data-driven analysis identifies four lung function trajectories. Trajectory membership has a genetic basis and is associated with distinct lung structural abnormalities.

Keywords: chronic obstructive pulmonary disease; lung function trajectories; longitudinal analysis

(Received in original form July 11, 2017; accepted in final form April 17, 2018)

*J.C.R. and P.J.C. contributed equally to this work.

‡G.R.W. and R.S.J.E. contributed equally to this work.

Supported by U.S. NIH grants R01 HL089856, R01 HL089897, R01 HL124233, R01 HL126596, and K25 HL130637. The NAS (Normative Aging Study) was supported by the Cooperative Studies Program-Epidemiology Research and Information Center of the U.S. Department of Veterans Affairs and is a component of the Massachusetts Veterans Epidemiology Research and Information Center (Boston, MA).

Author Contributions: Conception and design of this study and creation, revision, and final approval of this manuscript: J.C.R., P.J.C., M.H.C., C.P.H., F.N.R., G.V.S.-F., M.M.P., A.A.L., D.S., J.G.D., E.K.S., G.R.W., and R.S.J.E.; analysis and interpretation: J.C.R., P.J.C., G.V.S.-F., F.N.R., G.R.W., and R.S.J.E.; data acquisition: J.C.R., P.J.C., M.M.P., and R.S.J.E.; drafting the manuscript for important intellectual content: J.C.R., P.J.C., M.H.C., C.P.H., D.S., E.K.S., G.R.W., and R.S.J.E.

Correspondence and requests for reprints should be addressed to James C. Ross, Ph.D., Brigham and Women's Hospital, 45 Francis Street, Boston, MA 02115. E-mail: jross@bwh.harvard.edu.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Am J Respir Crit Care Med Vol 198, Iss 8, pp 1033–1042, Oct 15, 2018

Copyright © 2018 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201707-1405OC on April 19, 2018

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the

Subject: There are variable patterns of lung function decline, with low peak lung function attained in youth and periods of rapid lung function decline being major determinants of chronic obstructive pulmonary disease susceptibility.

What This Study Adds to the

Field: This study uses a data-driven approach that identifies four prototypical lung function trajectories in a longitudinal cohort of smokers. Analyzed on a separate, cross-sectional cohort, these trajectories have distinct phenotypic presentations. They are also differentially associated with self-reported parental history of chronic respiratory disease and genome-wide data-based estimates of heritability.

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death worldwide, and its prevalence continues to increase (1, 2). Clinically, the diagnosis and staging of COPD are based in large part on spirometric measures of lung function. An FEV₁/FVC ratio less than 0.7 is a threshold commonly used to detect the presence of expiratory airflow limitation. Decrements in the FEV₁ expressed as a percentage of predicted values are then used to grade disease severity. Although it was generally believed that smokers develop COPD primarily through an excessively rapid decline in lung function (3, 4), investigation has demonstrated that peak lung function attained in youth is a major determinant of disease susceptibility (5, 6), comorbidities, and mortality (7).

A challenge faced by many ongoing clinical, epidemiologic, and genetic investigations focused on smoking-related lung disease is that they lack historical data describing the trajectory of lung function decline from peak lung health to study inclusion. The lack of longitudinal perspective has also limited studies of COPD heterogeneity, which have primarily been performed using cross-sectional data (8). Two spirometrically indistinguishable smokers with moderate COPD may represent the extremes of low peak lung health and accelerated decline in lung

function, yet they are used interchangeably in attempts to understand the biologic basis of disease.

Although this may appear to be an insurmountable challenge for existing studies, we hypothesized that longitudinal data from unrelated but demographically similar cohorts could be used to define and model trajectories of FEV₁ decline. Those models could then be applied to cross-sectional studies or studies with limited observation time periods to assign participants to trajectories for further characterization of COPD heterogeneity.

To test these hypotheses, we used a data-driven approach called Bayesian nonparametric mixture modeling (9) to identify and characterize prototypical lung function trajectories, using longitudinal spirometry and smoking data from the NAS (Normative Aging Study). Using models of these trajectories, we then probabilistically assigned members of the more extensively phenotyped COPDGene (Genetic Epidemiology of COPD) Study cohort to each trajectory. We then assessed genetic and clinical differences between these trajectory subpopulations, using heritability analysis, comparisons of parental characteristics, and a comparative analysis of COPD-related measures including the BODE (body mass index, degree of airflow obstruction and dyspnea, and exercise capacity) index, MMRC (Modified Medical Research Council) dyspnea score, exercise capacity (6-min-walk distance), the number of COPD exacerbations over the previous year, and computed tomography (CT)-based measures of emphysema and airway disease.

Methods

The NAS is a longitudinal Veterans Administration study of healthy men established in 1963 (10). Men aged 21–80 years from the greater Boston area, free of known chronic medical conditions, were enrolled and underwent comprehensive clinical examinations at 5-year intervals for those less than 52 years old and at 3-year intervals for those more than 52 years old. Study data include spirometry performed in accordance with American Thoracic Society guidelines (11), completion of the American Thoracic Society Division of Lung Diseases 1978 questionnaire, and a separate questionnaire related to smoking

habits. Participants provided written informed consent at each visit, and the VA Boston Healthcare System Institutional Review Board approved the study.

The COPDGene Study is an ongoing multicenter, longitudinal study designed to investigate the genetic and epidemiologic characteristics of COPD (12). COPDGene enrolled 10,192 non-Hispanic white and African-American ever-smokers. Subjects were between the ages of 45 and 80 years and had a minimum 10-pack-year smoking history. Study data included the collection of detailed questionnaires, spirometric measures of lung function before and after the administration of short-acting bronchodilating medications, volumetric CT of the chest, and blood samples for genome-wide SNP genotyping. Genotyping was performed by Illumina on the HumanOmniExpress array. Subjects were excluded for missingness, heterozygosity, chromosomal aberrations, sex check, population outliers, and cryptic relatedness as previously described (13). A 5-year follow-up visit was performed, during which baseline data collection was repeated. The institutional review boards of all participating centers approved the COPDGene Study, and all participants provided written informed consent.

Chest CT analysis in the COPDGene Study has been described previously (12). Briefly, the Hounsfield unit (HU) value representing the 15th percentile of the lung region HU histogram (Perc15) was used for densitometric assessment of the lung parenchyma (14). Airway wall thickening was assessed as the square root of the wall area of a theoretical airway with an internal lumen perimeter of 10 mm (Pi10) (15). Finally, the percentage of lung tissue thought to represent functional small-airway disease (fSAD) was calculated, using parametric response mapping (PRM) as described previously (16).

PRM matches inspiratory and expiratory CT scans on a voxel-by-voxel basis to examine the change in density between images. By applying separate density thresholds to the inspiratory and expiratory voxel measurements, PRM discriminates between emphysema, nonemphysematous air trapping (fSAD), normal tissue, and “other” (image voxels not meeting the criteria for the other three categories). fSAD measures were transformed using the isometric log ratio transformation before utilization (17).

FEV₁ Trajectory Modeling

We assessed the existence of separate lung function trajectories in the Normative Aging Study by applying a Bayesian nonparametric trajectory mixture modeling approach (9) to those data. To faithfully apply models learned in NAS to COPDGene Study data, we restricted our analysis to those NAS subjects for whom at least one time point met COPDGene inclusion criteria for both age and smoking history (detailed in Reference 12 and summarized above). That is, for each NAS subject, at least one of their longitudinal time points had to correspond to an age between 45 and 80 years old with a corresponding smoking history of at least 10 pack-years. We also restricted our analysis in the COPDGene Study cohort to non-Hispanic white males with no more than 150 pack-years of smoke exposure (the approximate upper limit of exposure observed in NAS).

Our Bayesian-based identification of FEV₁ trajectories included as predictors age, height, smoking status, pack-years of tobacco exposure, and their higher-order terms (age², age³, etc.). The use of higher-

order terms (as opposed to just linear terms) enabled a nonlinear representation of FEV₁ decline: a trend that declines over time, but must stay above a lower limit to be compatible with life. The algorithm proceeds by iteratively refining the number and shape of trajectories as well as the assignment of subjects to trajectories. When no further improvement in data fit is detected, the algorithm terminates and returns probabilistic models of each trajectory. Using the Watanabe-Akaike information criterion (18), we selected a set of FEV₁ predictors that gave both an accurate and parsimonious description of the data. This process led to the identification of four spirometric trajectories.

Next, we used these trajectory models to assign each COPDGene subject to the most probable trajectory, using baseline data as well as follow-up data for those subjects who had both. For example, supposing a COPDGene subject had a 0.8 probability of belonging to trajectory 1, a 0.1 probability of belonging to trajectory 2, and a 0.05 probability of belonging to trajectories 3 and 4, we would assign the subject to trajectory 1 (*see* the online

supplement for details). Note that the use of follow-up (in addition to baseline) data for the COPDGene subjects affects the probability of trajectory membership. To assess the effect of using both baseline and follow-up data as opposed to baseline data alone, we considered the 1,802 COPDGene subjects for whom we have both and computed trajectory assignments using 1) baseline data only and 2) both baseline and follow-up data. In addition, for all COPDGene subjects assigned to each of the trajectories, we computed the average probability of assignment within each trajectory group.

Statistical Analysis

We stratified COPDGene Study baseline data by age (45–55, 55–65, 65–75, and 75–85 yr old) and compared quantitative CT characteristics, lung function, and 6-minute-walk distance between trajectories within age strata. Statistical comparisons were performed by Welch's unequal variances *t* test (19) implemented in *scipy.stats* software (20) (version 0.17.0). The BODE index (21), MMRC dyspnea score, and number of COPD

Table 1. Characteristics of Subjects Included in Study

Characteristic	Normative Aging Study		COPDGene Study	
	Baseline	Last Visit	Baseline	5-Year Follow-up
<i>n</i>	1,060	1,060	3,546	1,802
Visit No.	1	6.3 ± 2.4	1	2
Age, yr				
Mean ± SD	42.2 ± 8.6	68.9 ± 8.6	62.3 ± 8.8	68.0 ± 8.2
Range	24.7–77.2	45.1–91.6	45.0–81.0	47.0–86.8
FEV ₁				
Mean ± SD, L	3.8 ± 0.6	2.6 ± 0.7	2.5 ± 1.0	2.4 ± 0.9
Range, L	0.7–6.1	0.7–4.9	0.3–5.5	0.4–5.1
% of predicted value	116.2 ± 15.8	104.8 ± 32.1	72.6 ± 26.4	75.2 ± 25.6
FVC				
Mean ± SD, L	4.8 ± 0.7	3.6 ± 0.7	3.9 ± 1.0	3.7 ± 0.9
Range, L	2.5–6.6	1.6–6.3	1.2–7.7	1.2–7.1
% of predicted value	88.2 ± 10.8	93.0 ± 14.5	85.3 ± 18.2	85.7 ± 17.8
FEV ₁ /FVC, %				
Mean ± SD	79 ± 7	72 ± 21	63 ± 17	64 ± 16
Range	17.6–98.8	17.6–92.8	15.0–94.0	17.0–92.0
BMI, kg/m ²				
Mean ± SD	25.7 ± 2.9	27.9 ± 4.0	28.8 ± 5.5	29.1 ± 5.5
Range	16.0–38.4	15.8–52.3	13.8–58.6	16.7–58.6
Pack-years				
Mean ± SD	29.0 ± 17.9	40.3 ± 24.0	50.6 ± 25.8	49.6 ± 24.8
Range	0.0–117.0	10.0–145.4	10.0–150.0	10.0–146.8
Height, cm				
Mean ± SD	175.9 ± 6.4	173.2 ± 6.5	176.1 ± 6.9	175.7 ± 6.8
Range	153.9–194.1	148.5–191.3	138.9–200.3	152.0–199.4
Current smoker, No./total No. (%)	603/1,060 (57)	188/1,060 (18)	1,427/3,546 (40)	471/1,802 (26)

Definition of abbreviations: BMI = body mass index; COPDGene = Genetic Epidemiology of COPD Study.

Figure 1. Lung function trajectories identified by nonparametric Bayesian mixture modeling in the Normative Aging Study. Connected dots indicate sequences of subject visits. Thick lines indicate expected (population-averaged) FEV₁ values as a function of age (assuming average height and pack-years of smoke exposure). Shaded regions represent FEV₁ 95% prediction interval estimated by the mixture modeling approach (see DISCUSSION OF PREDICTION INTERVALS in the online supplement for additional details).

Table 2. Characteristics of COPDGene Subjects by Trajectory Assignment Stratified by Age*

	45–55 Years Old				55–65 Years Old			
	T1	T2	T3	T4	T1	T2	T3	T4
<i>n</i>	98	272	361	96	243	411	545	82
FEV ₁ , % pred	41.1 ± 10	73.3 ± 10	94.4 ± 8	110.2 ± 7	34 ± 10	66.2 ± 10	91.8 ± 10	112 ± 10
FEV ₁ , L	1.60 ± 0.59	2.82 ± 0.49	3.68 ± 0.43	4.45 ± 0.38	1.21 ± 0.46	2.33 ± 0.55	3.24 ± 0.42	3.93 ± 0.40
FVC, % pred	69 ± 20	83.5 ± 10	95.9 ± 10	108.4 ± 8	65.8 ± 20	80.7 ± 10	95 ± 10	111.1 ± 10
FVC, L	3.46 ± 1.0	4.13 ± 0.8	4.82 ± 0.7	5.63 ± 0.6	3.07 ± 0.9	3.75 ± 0.8	4.44 ± 0.6	5.17 ± 0.6
FEV ₁ /FVC	0.47 ± 0.14	0.69 ± 0.10	0.77 ± 0.06	0.79 ± 0.05	0.40 ± 0.12	0.63 ± 0.13	0.74 ± 0.07	0.76 ± 0.06
Height, cm	177.2 ± 7	176.1 ± 7	177.3 ± 7	179.7 ± 6	175.9 ± 7	176 ± 7	176.4 ± 7	176.2 ± 7
BMI, kg/m ²	27.8 ± 6	29.4 ± 7	28.2 ± 5	26.8 ± 4	28.1 ± 6	29.9 ± 6	28.9 ± 5	28.2 ± 4
Pack-years	44.3 ± 20	40.5 ± 20	38 ± 20	41.4 ± 20	52.2 ± 20	54.3 ± 30	46.4 ± 20	53.6 ± 30
Pi10	3.75 ± 0.2	3.68 ± 0.1	3.61 ± 0.1	3.58 ± 0.1	3.74 ± 0.2	3.66 ± 0.1	3.59 ± 0.1	3.56 ± 0.1
Perc15, HU	−939.7 ± 30	−910 ± 20	−912.4 ± 20	−919.9 ± 20	−950.3 ± 30	−922.1 ± 30	−920 ± 20	−922.9 ± 20
BODE	3.73	1.06	0.36	0.28	4.24	1.40	0.28	0.22
MMRC	2.42	1.32	0.65	0.48	2.63	1.41	0.52	0.44
Exacerbations	0.91	0.35	0.10	0.05	0.97	0.45	0.10	0.06
6MWD, m	378 ± 122	456 ± 110	492 ± 103	511 ± 98	358 ± 117	427 ± 115	490 ± 107	502 ± 109

	65–75 Years Old				75–85 Years Old			
	T1	T2	T3	T4	T1	T2	T3	T4
<i>n</i>	235	459	412	46	39	141	93	13
FEV ₁ , % pred	30.6 ± 10	55.9 ± 10	89.4 ± 10	111.5 ± 10	32.9 ± 8	57.5 ± 20	90.7 ± 10	114.3 ± 7
FEV ₁ , L	0.99 ± 0.36	1.76 ± 0.54	2.84 ± 0.45	3.76 ± 0.45	0.95 ± 0.27	1.59 ± 0.47	2.53 ± 0.43	3.31 ± 0.28
FVC, % pred	64.3 ± 10	76.8 ± 10	93.7 ± 10	110 ± 10	69 ± 20	80.3 ± 20	93.7 ± 10	107.4 ± 8
FVC, L	2.82 ± 0.7	3.26 ± 0.7	4.04 ± 0.6	5.02 ± 0.6	2.76 ± 0.8	3.10 ± 0.6	3.65 ± 0.6	4.35 ± 0.4
FEV ₁ /FVC	0.35 ± 0.10	0.54 ± 0.14	0.71 ± 0.08	0.75 ± 0.06	0.35 ± 0.08	0.52 ± 0.13	0.70 ± 0.10	0.77 ± 0.05
Height, cm	177 ± 7	174.7 ± 7	175.7 ± 7	179.1 ± 7	175.7 ± 6	173.8 ± 7	174.5 ± 7	176.8 ± 5
BMI, kg/m ²	27.2 ± 5	29.5 ± 6	29.4 ± 5	28 ± 4	26.9 ± 5	28.5 ± 5	28.9 ± 5	29.4 ± 5
Pack-years	52 ± 20	63.3 ± 30	55.3 ± 30	55.2 ± 20	42.1 ± 20	54.2 ± 30	54.1 ± 30	49.7 ± 20
Pi10	3.76 ± 0.1	3.68 ± 0.1	3.59 ± 0.1	3.57 ± 0.1	3.78 ± 0.2	3.70 ± 0.2	3.62 ± 0.1	3.54 ± 0.1
Perc15, HU	−959.5 ± 20	−936.6 ± 30	−925.2 ± 20	−930.6 ± 10	−962.6 ± 20	−940 ± 20	−927.8 ± 20	−930.1 ± 10
BODE	4.57	2.16	0.31	0.04	4.59	1.98	0.22	0.15
MMRC	2.65	1.59	0.61	0.15	2.74	1.57	0.51	0.31
Exacerbations	0.76	0.52	0.15	0.07	0.82	0.49	0.29	0.00
6MWD, m	301 ± 115	399 ± 113	461 ± 101	512 ± 87	301 ± 109	365 ± 100	415 ± 98	459 ± 96

Definition of abbreviations: % pred = percentage of the predicted value; 6MWD = 6-minute-walk distance; BMI = body mass index; BODE = body mass index, airflow obstruction, dyspnea, and exercise capacity index; COPDGene = Genetic Epidemiology of COPD Study; HU = Hounsfield units; MMRC = modified Medical Research Council; Perc15 = HU value representing the 15th percentile of the lung region HU histogram; Pi10 = internal lumen perimeter of 10 mm.

*Values represent means ± SD.

exacerbations over the previous year were related to trajectory membership using ordinal logistic regression (R version 3.1 [22]).

We used Pearson's χ^2 test (23) (scipy.stats version 0.17.0 [20]) to assess associations to presence of maternal and paternal asthma, emphysema, chronic bronchitis, and COPD. Trajectory heritability was estimated in COPDGene using methods developed from genome-wide genotyping data from unrelated population-based samples (24). In this method, narrow sense heritability quantifies the proportion of phenotypic variance that can be explained by genetic variance under an additive genetic model. The genetic similarity matrix was estimated from 664,892 genotyped, autosomal SNPs with a minor allele frequency greater than 0.01 in all of the non-Hispanic white subjects

and a Hardy-Weinberg equilibrium *P* value less than 10^{-8} with available data from the COPDGene Study ($n = 6,678$) as previously described (25). Using a subset of the genetic similarity matrix corresponding to the subjects analyzed for this analysis, heritability was calculated for each of the six possible contrasts between the four trajectories, using the restricted maximum likelihood method implemented in the GCTA software package (version 1.13) (24).

Results

A total of 1,060 NAS participants met COPDGene inclusion criteria on at least one study visit. The median follow-up time for these subjects was 29 years. The COPDGene subcohort consisted of 3,546 non-Hispanic

white males after excluding those who had more than 150 pack-years of tobacco smoke exposure. Of those, 1,802 have returned for their 5-year follow-up visit as of September 24, 2016. NAS subjects were on average 20 years younger at baseline and 6.5 years older on their final visit than COPDGene subjects at baseline. NAS subjects also had significantly less tobacco smoke exposure (as reported on their final study visit) than smokers in COPDGene (Table 1).

Nonparametric Bayesian mixture modeling identified four lung function trajectories in the Normative Aging Study (Figure 1 and the online supplement). Trajectory 1 appeared to represent those with the lowest peak lung health and the most rapid decline in lung function, whereas trajectories 2, 3, and 4 represented subjects with incrementally greater peak lung health

Figure 2. Airway wall thickening (left) and emphysema (right) trends in COPDGen (Genetic Epidemiology of COPD Study) by trajectory assignment, stratified by age. Airway wall thickening is assessed as the square root of the wall area ($\sqrt{WA}$) of a theoretical airway with an internal lumen perimeter of 10 mm (14). Emphysema is assessed as the Hounsfield unit (HU) value representing the 15th percentile of the lung region HU histogram (Perc15) (13). Greater amounts of emphysema are seen in the trajectory 2 subgroup relative to trajectories 3 and 4 only within older age strata, and thicker airway walls are observed in trajectory 2 individuals throughout ($P < 0.001$ for all airway thickness comparisons).

but relatively similar rates of decline in lung function. Factors found to give the best fit to these data were as follows: age, age², age³, age⁴, height², pack-years, pack-years², pack-years³, pack-years⁴, and an intercept term. Statistical comparison of rates of decline between trajectories and association with Global Initiative for Chronic Obstructive Lung Disease spirometric stage can be found in the section TRAJECTORY ANALYSIS IN THE NORMATIVE AGING STUDY in the online supplement.

As stated earlier, in the section FEV₁ TRAJECTORY MODELING, we assessed the effect of using both baseline and follow-up data as opposed to baseline data alone when assigning COPDGen subjects to trajectories. For the 1,802 COPDGen subjects for whom we had both baseline and follow-up data, we observed that trajectory assignment changed for 243 (13%) subjects. The breakdown of trajectory reassignment is given in Table E2. In addition, Table E3 gives the average probability of assignment within each trajectory group (assuming the use of follow-up data). Table E3 indicates that, on average, trajectory assignments are made with high probability. The following results assume COPDGen trajectory assignments made with baseline and available follow-up data.

Table 2 shows characteristics of COPDGen subjects by trajectory assignment and age strata. Trajectory 1 subjects were found to have the greatest radiologic burden of disease, with the most emphysema, thicker airway walls, and more

functional small-airway disease across all age strata. All results were significant at the $P < 0.01$ significance level. Trajectory 2 subjects had thicker airway walls than those assigned to trajectories 3 and 4 at all age strata (Figure 2) as well as more emphysema and fSAD with increasing age (Figure 3). Subjects in trajectories 3 and 4 did not develop appreciable emphysema on CT despite averaging more than 45 pack-years of tobacco exposure. In the highest age strata, subjects in trajectory 3 had thicker airway walls than their counterparts in trajectory 4. Overall, membership in lower lung function trajectories was associated with worse exercise capacity, higher BODE scores, greater dyspnea, and more frequent exacerbations over the previous year (see Tables E4, E7, E10, and E13).

Parental characteristics for each trajectory and statistical comparisons between trajectories are provided in Table 3. Subjects in trajectories 1 and 2 had greater self-reported parental histories of emphysema, COPD, chronic bronchitis, and asthma than did subjects in trajectories 3 and 4.

Using genome-wide genotyping data in COPDGen, we estimated how much of the variability in trajectory assignment could be explained by genetic differences, and we found that the estimated heritability of trajectory assignment was high for some trajectories. Of the six possible pairwise comparisons of the four trajectories, four comparisons had statistically significant evidence of a genetic contribution.

The genetic contribution to trajectory 1 was particularly high, as the estimated heritabilities for this trajectory, compared with trajectories 3 and 4, were 51% ($P = 0.003$) and 83% ($P = 0.02$), respectively.

Discussion

Using an approach called Bayesian nonparametric mixture modeling, we leveraged separate but complementary cohorts to better understand patterns of lung function decline. Models of lung function decline learned from longitudinal data in the Normative Aging Study were applied to baseline data (and 5-yr follow-up data for some subjects) in the COPDGen Study. This approach provided novel insights into the natural history of COPD and COPD phenotypic heterogeneity.

In particular, we identified four prototypical lung function trajectories in the Normative Aging Study, demonstrating that the concept of multiple trajectories (5, 6) is well supported by data-driven analysis of empirical data. The analysis of trajectories in the COPDGen Study provides novel insight into the temporal relationship between distinct phenotypic aspects of COPD within trajectories. We observed significantly more airway wall thickening (Pi10) in the trajectory 2 subgroup compared with the trajectory 3 and 4 subgroups across all age strata. However, we see evidence of increased emphysema

Figure 3. Age-stratified average relative composition of emphysema (Emph), functional small-airway disease (fSAD), normal tissue, and “other” (dark yellow portions) for each trajectory, as assessed by parametric response mapping analysis (12). Rows correspond to age strata, and columns correspond to trajectories; color-coded arrows clarify the direction of increasing age for each trajectory.

Table 3. Heritability and Family History of Respiratory Disease in COPDGene Subjects Stratified by Lung Function Trajectory

Characteristic*	T1 (n = 615 [17%])	T2 (n = 1,283 [36%])	T3 (n = 1,411 [40%])	T4 (n = 237 [7%])	P Value†					
					T1 vs. T2	T1 vs. T3	T1 vs. T4	T2 vs. T3	T2 vs. T4	T3 vs. T4
Father, emphysema	127/333 (27.6%)	192/764 (20.1%)	174/937 (15.7%)	28/150 (15.7%)	0.002	<0.001	0.002	0.01	0.21	0.93
Mother, emphysema	87/411 (17.5%)	124/914 (11.9%)	123/1,078 (10.2%)	10/188 (5.1%)	0.004	<0.001	<0.001	0.22	0.006	0.03
Father, COPD	75/364 (17.1%)	106/802 (11.7%)	87/968 (8.2%)	12/150 (7.4%)	0.008	<0.001	0.004	0.01	0.14	0.83
Mother, COPD	74/424 (14.9%)	79/920 (7.9%)	82/1,086 (7.0%)	8/186 (4.1%)	<0.001	<0.001	<0.001	0.48	0.09	0.18
Father, chronic bronchitis	44/379 (10.4%)	70/820 (7.9%)	61/1,002 (5.7%)	7/167 (4.0%)	0.16	0.002	0.02	0.08	0.1	0.46
Mother, chronic bronchitis	45/440 (9.3%)	87/926 (8.6%)	94/1,088 (8.0%)	10/188 (5.1%)	0.73	0.43	0.09	0.64	0.12	0.2
Father, asthma	29/412 (6.6%)	44/880 (4.8%)	43/1,062 (3.9%)	6/172 (3.4%)	0.21	0.03	0.17	0.39	0.54	0.9
Mother, asthma	30/454 (6.2%)	75/942 (7.4%)	51/1,135 (4.3%)	8/189 (4.1%)	0.47	0.13	0.36	0.003	0.13	0.97
Heritability‡	0.02 (0.14 ± 0.18) 0.003 (0.51 ± 0.18) 0.02 (0.83 ± 0.40) 0.2 (0.10 ± 0.12) 0.2 (0.17 ± 0.22) 0.04 (0.33 ± 0.20)									

Definitions of abbreviations: COPD = chronic obstructive pulmonary disease; COPDGene = Genetic Epidemiology of COPD Study.

*Proportions given are of those with versus those without (with/without). Percentages indicate the percentage of those with relative to all respondents [100 × with/(with + without)].

†P values are based on Pearson's χ^2 test.

‡Values in parentheses: narrow sense heritability (quantifies the proportion of phenotypic variance that can be explained by genetic variance under an additive genetic model).

(Perc15) only in the trajectory 2 subgroup compared with the trajectory 3 and 4 subgroups in older age strata: 65–75 and 75–85 years (see Tables E4, E7, E10, E13, and Figure 2). This suggests that subjects in the airway/emphysema trajectory (trajectory 2) develop airway wall thickening in advance of emphysema (under the assumption that COPDGene subjects of different ages who have been assigned to the same trajectory represent different stages of progression along that trajectory). In contrast, subjects in the airway-predominant trajectory (trajectory 3) have isolated airway wall thickening and do not develop emphysema to the same degree throughout the entire observation period. The observation that, for some subjects, airway disease precedes emphysema is in accordance with the pathologic studies of Hogg and colleagues demonstrating that small-airway wall inflammation and dropout precede the development of emphysema (26). Although one might hypothesize that the same pattern also applies to the rapid, early decline trajectory (trajectory 1), the majority of this group already had established, severe obstruction before enrollment in COPDGene, and thus their CT imaging data do not capture the proper time interval to observe this phenomenon.

The trajectories identified in our study were also differentially associated with self-reported parental history of chronic respiratory disease and genome-wide data-based estimates of heritability, providing further evidence for the genetic underpinnings of COPD heterogeneity. Heritability estimates for COPD have typically been in the range of 30%, indicating that this proportion of the overall variability in developing COPD is due to genetic factors (25). The heritability estimates obtained for pairwise

comparisons of trajectories in this study exceeded 50% for the comparisons between trajectory 1 and trajectories 3 and 4, suggesting that most of the phenotypic differences between the rapid early decline trajectory and the trajectories with more preserved lung function are due to underlying genetic differences between the individuals in these trajectories. Although these results should be confirmed in other cohorts, they are consistent with the intuition that genetic differences between groups of subjects will be more observable once confounding factors of age and lifetime smoke exposure are removed. Trajectory-based analyses of COPD, such as the method presented herein, mitigate these confounding effects by placing individuals within the context of longitudinal decline rather than analyzing a single cross-section of subjects at various stages in their disease course.

This study has several important limitations. The NAS is limited to males (mostly veterans), so our current findings are limited to males and not necessarily generalizable to the whole population. Future studies of cohorts with adequate follow-up in female subjects are needed. COPDGene is a convenience sample of smokers enriched for COPD, and subjects were all 45 years or older at baseline. The longitudinal follow-up in COPDGene is approximately 5 years, so these data alone are not sufficient to describe lifelong lung function trajectories. To address this problem, we learned longitudinal trajectories from NAS, and then COPDGene subjects were assigned to these trajectories using their baseline and follow-up data. Although the model has good performance characteristics, it would be preferable to have more extensive follow-up data in COPDGene to be able to directly observe the lung function in these subjects before age 45. In addition, direct observation of CT

image data in subjects before the age of 45 would provide further insight into this critically important period of lung function decline. NAS and COPDGene differ in their enrollment criteria, so the studies in their entirety are not directly comparable. To account for these differences, we limited the NAS analysis to subjects who would have met the COPDGene enrollment criteria for at least one of their observation time points, and we limited the COPDGene analysis to males only. Finally, our Bayesian model does not incorporate mortality data, and it is possible that our results suffer from survival bias. The “flattening out” of the trajectory 1 curve seen in Figure 1 is particularly suggestive of this phenomenon, as it is more likely for those with low peak lung function early in life to be at greater risk for all-cause mortality relative to those with higher peak lung function (7). In the future, we plan to augment our Bayesian algorithm to jointly consider time-to-event (mortality) data as well as longitudinal data for a more robust definition of trajectories.

In summary, the Bayesian modeling approach allows for the combined analysis of two cohorts with complementary data, one which was well suited for learning lung function trajectory data from early adulthood to late life and a second cohort that contained extensive radiologic and genetic information obtained over a 5-year period of observation. These trajectories were associated with different amounts of risk of developing COPD and distinct COPD-related phenotypic manifestations. Analysis of these patterns confirms that decreased levels of maximally attained FEV₁ and periods of rapid lung function decline both contribute to COPD risk, and heritability analyses indicate that lung function trajectories are in part genetically determined. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

1. World Health Organization. World health statistics 2008. Geneva, Switzerland: World Health Organization; 2008.
2. Minino AM, Murphy SL. Death in the United States, 2010. *NCHS Data Brief* 2012;99:1–8.
3. Fletcher C, Peto R. The natural history of chronic airflow obstruction. *Br Med J* 1977;1:1645.
4. Speizer FE, Tager IB. Epidemiology of chronic mucus hypersecretion and obstructive airways disease. *Epidemiol Rev* 1979;1:124–142.
5. Lange P, Celli B, Agustí A, Boje Jensen G, Divo M, Faner R, et al. Lung-function trajectories leading to chronic obstructive pulmonary disease. *N Engl J Med* 2015;373:111–122.
6. McGeachie MJ, Yates KP, Zhou X, Guo F, Sternberg AL, Van Natta ML, et al. Patterns of growth and decline in lung function in persistent childhood asthma. *N Engl J Med* 2016;374:1842–1852.
7. Agustí A, Noell G, Brugada J, Faner R. Lung function in early adulthood and health in later life: a transgenerational cohort analysis. *Lancet Respir Med* 2017;5:935–945.
8. Pinto LM, Alghamdi M, Benedetti A, Zaihra T, Landry T, Bourbeau J. Derivation and validation of clinical phenotypes for COPD: a systematic review. *Respir Res* 2015;16:50.
9. Ross JC, Castaldi PJ, Cho MH, Chen J, Chang Y, Dy JG, et al. A Bayesian nonparametric model for disease subtyping: application to emphysema phenotypes. *IEEE Trans Med Imaging* 2017;36:343–354.

10. Bell B, Rose CL, Damon A. The Normative Aging Study: an interdisciplinary and longitudinal study of health and aging. *Int J Aging Hum Dev* 1972;3: 5–17.
11. Sparrow D, O'Connor G, Colton T, Barry CL, Weiss ST. The relationship of nonspecific bronchial responsiveness to the occurrence of respiratory symptoms and decreased levels of pulmonary function: the Normative Aging Study. *Am Rev Respir Dis* 1987;135:1255–1260.
12. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD* 2010;7:32–43.
13. Cho MH, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, et al.; NETT Genetics, ICGN, ECLIPSE, and COPDGene Investigators. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* 2014;2: 214–225.
14. Gould GA, MacNee W, McLean A, Warren PM, Redpath A, Best JJ, et al. CT measurements of lung density in life can quantitate distal airspace enlargement: an essential defining feature of human emphysema. *Am Rev Respir Dis* 1988;137:380–392.
15. Nakano Y, Wong JC, de Jong PA, Buzatu L, Nagao T, Coxson HO, et al. The prediction of small airway dimensions using computed tomography. *Am J Respir Crit Care Med* 2005;171:142–146.
16. Galbán CJ, Han MK, Boes JL, Chughtai KA, Meyer CR, Johnson TD, et al. Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat Med* 2012;18:1711–1715.
17. Egozcue JJ, Pawłowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C. Isometric logratio transformations for compositional data analysis. *Math Geol* 2003;35:279–300.
18. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Mach Learn Res* 2010;11:3571–3594.
19. Welch BL. The generalisation of student's problems when several different population variances are involved. *Biometrika* 1947;34:28–35.
20. Jones E, Oliphant E, Peterson P, et al. SciPy: open source scientific tools for Python, 2001– [accessed 21 June 2018]. Available from: <http://www.scipy.org/>.
21. Celli BR, Cote CG, Marin JM, Casanova C, Montes de Oca M, Mendez RA, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med* 2004;350:1005–1012.
22. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016. Available from: <https://www.R-project.org/>.
23. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Lond Edinb Dublin Philos Mag J Sci* 1900;50:157–175.
24. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;88:76–82.
25. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *Am J Respir Crit Care Med* 2013;188:941–947.
26. Hogg JC, Chu F, Utokaparch S, Woods R, Elliott WM, Buzatu L, et al. The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N Engl J Med* 2004;350:2645–2653.