ORIGINAL ARTICLE

Phenotypes of Rapid Cystic Fibrosis Lung Disease Progression during Adolescence and Young Adulthood

Rhonda D. Szczesniak[1,2,3], Dan Li[4], Weiji Su[5], Cole Brokamp[1], John Pestian[3,6], Michael Seid[2,3,7], and John P. Clancy[2,3]

[1]Division of Biostatistics and Epidemiology, [2]Division of Pulmonary Medicine, [6]Division of Biomedical Informatics, and [7]James M. Anderson Center for Health Systems Excellence, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; [3]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio; [4]Alzheimer's Therapeutic Research Institute, Keck School of Medicine, University of Southern California, Los Angeles, California; and [5]Department of Mathematical Sciences, University of Cincinnati, Cincinnati, Ohio

Abstract

Rationale: Individuals with cystic fibrosis are at risk for prolonged drops in lung function, clinically termed rapid decline, during discreet periods of the disease.

Objectives: To identify phenotypes of rapid pulmonary decline and determine how these phenotypes are related to patient characteristics.

Methods: A longitudinal cohort study of patients with cystic fibrosis aged 6–21 years was conducted using the Cystic Fibrosis Foundation Patient Registry. A statistical approach for clustering longitudinal profiles, sparse functional principal components analysis, was used to classify patients into distinct phenotypes by evaluating trajectories of FEV_1 decline. Phenotypes were compared with respect to baseline and mortality characteristics.

Measurements and Main Results: Three distinct phenotypes of rapid decline were identified, corresponding to early, middle, and late timing of maximal FEV_1 loss, in the overall cohort (n = 18,387). The majority of variation (first functional principal component, 94%) among patient profiles was characterized by differences in mean longitudinal FEV_1 trajectories. Average degree of rapid decline was similar among phenotypes (roughly −3% predicted/yr); however, average timing differed, with early, middle, and late phenotypes experiencing rapid decline at 12.9, 16.3, and 18.5 years of age, respectively. Individuals with the late phenotype had the highest initial FEV_1 but experienced the greatest loss of lung function. The early phenotype was more likely to have respiratory infections and acute exacerbations at baseline or to develop them subsequently, compared with other phenotypes.

Conclusions: By identifying phenotypes and associated risk factors, timing of interventions may be more precisely targeted for subgroups at highest risk of lung function loss.

Keywords: cluster analysis; epidemiology; functional data analysis; nonlinear trajectories; pulmonary function

At a Glance Commentary

Scientific Knowledge on the Subject: Rapid lung function decline during adolescence and early adulthood is a frequent observation in the clinical course of patients with cystic fibrosis. However, the heterogeneity in the timing of rapid lung function decline remains unknown.

What This Study Adds to the Field: Using a novel statistical approach to model lung function trajectories, this study suggests that phenotypic variation in the clinical course of cystic fibrosis is more strongly linked to the timing, as opposed to extent, of rapid decline.

(Received in original form December 28, 2016; accepted in final form April 13, 2017)

Supported by NHLBI (National Institutes of Health) grant K25 HL125954 and Cystic Fibrosis Foundation Research and Development Program grant R457-CR11.

Author Contributions: Conception and design, R.D.S. and J.P.C. Analysis and interpretation, R.D.S., D.L., W.S., C.B., and J.P.C. Intellectual revision and drafting of the manuscript, and approval of the final version, R.D.S., D.L., W.S., C.B., J.P., M.S., and J.P.C.

Correspondence and requests for reprints should be addressed to Rhonda D. Szczesniak, Ph.D., Division of Biostatistics and Epidemiology (MLC 5041), Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229. E-mail: rhonda.szczesniak@cchmc.org

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 196, Iss 4, pp 471–478, Aug 15, 2017

Originally Published in Press as DOI: 10.1164/rccm.201612-2574OC on April 14, 2017
Internet address: www.atsjournals.org

Slowing lung disease progression is essential to survival in individuals with cystic fibrosis (CF). Decline in pulmonary function, measured as percentage predicted FEV_1, is the strongest predictor of mortality in this population (1). However, there is substantial variation within each individual patient's FEV_1 trajectory and between patient trajectories observed over time (2); this variation is evident on a week-to-week basis (3). Many patient characteristics have been linked to faster decline in lung function (4, 5), but beyond the slope, there has been limited information extracted from the inherent shape of the FEV_1 trajectory.

Recent studies using flexible modeling approaches that allow for analysis of nonlinear FEV_1 trajectories have more precisely characterized the timing of rapid decline as occurring predominantly during adolescence and early adulthood. In a retrospective study using the U.S. Cystic Fibrosis Foundation Patient Registry (CFFPR), we found that the timing and degree of rapid FEV_1 decline were variable between patients (6). Individuals experienced most rapid decline at a median (interquartile range) of 16.3 (13.5–21.0) years and the average degree of maximal FEV_1 loss was 2.0% predicted/yr. In another recent retrospective single-center analysis, Moss and colleagues (7) used a random change point model to fit FEV_1 trajectories of patients with CF, allowing the timing of rapid decline to differ by individual. In addition to variable decline, each individual FEV_1 trajectory exhibits strong within-patient correlation. A study of the Danish CF registry using monthly data from 1969 to 2010 confirmed that individual FEV_1 observations taken as much as 15 years apart remain correlated (2).

Although these epidemiologic studies have begun to shed light on the complex nature of FEV_1 decline by using mixed effects models, clustering approaches are needed to establish phenotypes of rapid decline and appraise their clinical relevance in CF. Clustering curve data with strong temporal correlation, like that of longitudinal FEV_1 in CF, can be accomplished using a functional data analysis technique known as functional principal components analysis for sparse longitudinal data (FPCA) (8). This model allows individual prediction of smoothed curves of nonlinear FEV_1 decline. It also accounts for temporal correlation and use of profiles with varying numbers of FEV_1 values. Thus, this methodology is well suited to CFFPR analyses, where numbers of values and trajectories of FEV_1 vary over time.

A better characterization of distinct patterns or phenotypes of rapid lung function decline would provide an opportunity for more precise, personalized treatment of patients with CF. Thus, the purposes of this study were threefold: first, to identify and characterize phenotypes of rapid FEV_1 decline for adolescents and young adults with CF; second, to identify predictors of earlier rapid FEV_1 decline; and third, to determine the extent to which other longitudinal characteristics are associated with rapid FEV_1 decline phenotypes. A portion of the results of this study has been previously reported in the form of an abstract (9).

Methods

Population

Patients with FEV_1 data recorded during childhood and adolescence (age, 6–21 yr) in the U.S. CFFPR between January 1, 1997, and December 31, 2013, were eligible. Patients younger than 6 years of age were excluded because of potentially unreliable pulmonary function testing. Because the majority of relevant predictors of FEV_1 decline were consistently documented in the CFFPR beginning in 1997, we considered available data from 1997 onward. A detailed description of this registry and its contents is provided elsewhere (10).

Design and Procedures

We performed a retrospective longitudinal cohort study using the CFFPR, clustering patients with CF into phenotypes according to patterns of age-related lung function decline. Observed FEV_1 was expressed as percentage of predicted using established reference equations (11, 12). Baseline was defined as the time at which the first FEV_1 was recorded during the eligibility period. Data acquired after lung transplant were excluded. The institutional review board at Cincinnati Children's Hospital Medical Center approved the study.

Statistical Analysis

Descriptive statistics, including mean (SD) or median (range) as appropriate for continuous variables and number (%) for categorical variables, were used to summarize cohort characteristics. Chi-square tests or type 3 tests of fixed effects from a linear model, as appropriate, were used to examine overall differences in phenotype groups with respect to each variable. Details of the statistical considerations and implementation code are available in the online supplement.

Longitudinal pattern classification. FEV_1 trajectories were classified using FPCA for sparse longitudinal data. It is similar to principal components analysis, which finds linear combinations of a small number of features to maximize variance across data. By looking at these features as functional data, FPCA can extract the common temporal characteristics of a set of curves. The longitudinal outcome variable used in the FPCA was mean quarterly FEV_1. Before conducting FPCA, we fit each patient's FEV_1 trajectory based on our previous approach (13). For inclusion, each patient had to contribute at least seven FEV_1 measurements over time, to fit individual curves used as the unit of information in the FPCA to identify distinct patterns (*see* Section E1 in the online supplement). The number of functional principal components (FPCs) retained in FPCA was based on the standard eigenvalue criterion ($\geq$80% of variance explained) (14).

Given the amount of variation explained by the first FPC, we followed a previously described approach (15) to classify the fitted curves into three clusters using the first and third quartiles (Q1 and Q3, respectively) of scores from the first FPC. Patients with scores less than Q1 were classified into the first cluster; patients with scores between Q1 and Q3 were classified into the second cluster; patients with scores greater than Q3 were classified into the third cluster. Further details are provided in Section E2. Rate of FEV_1 decline for each phenotype was estimated by taking the derivative over age corresponding to the fitted FEV_1 trajectory; the approach is detailed in Section E3. Sensitivity analyses, also provided as supplemental material

(*see* Section E4), were performed to assess potential selection bias arising from restricting FPCA to data from patients with at least seven FEV_1 measurements, and to examine the extent to which FPCA findings may be impacted by variation in the length of follow-up across patients. Analyses were implemented using R 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria).

Modeling predictors of phenotype with early decline. To identify characteristics predictive of the early phenotype of rapid lung function decline obtained from FPCA, we fit a logistic regression model, with early decline phenotype membership (Yes or No) as the primary outcome. Baseline predictors that were included in the model as covariates were birth cohort, age at first FEV_1, age at CF diagnosis, FEV_1% predicted, body mass index (BMI) percentile, pancreatic enzyme use, number of acute exacerbations within the year, infection with methicillin-resistant *Staphylococcus aureus* (MRSA), *Pseudomonas aeruginosa*, *Burkholderia cepacia*, nontuberculous mycobacteria (NTM), *Stenotrophomonas*, allergic

Table 1. Baseline and Mortality Characteristics of Patients with Cystic Fibrosis, Overall and by Phenotypes of Lung Function Decline

Patient Characteristics	Overall	Phenotype*			*P* Value
		Early	Middle	Late	
N (%)	18,387 (100)	4,597 (25.0)	9,194 (50.0)	4,596 (25.0)	
Men	9,524 (51.8)	2,142 (46.6)	4,854 (52.8)	2,504 (54.5)	<0.0001
Δ F508 copies					<0.0001
None	3,806 (20.7)	1,195 (26.0)	1,811 (19.7)	822 (17.9)	
1	8,568 (46.6)	2,055 (44.7)	4,321 (47.0)	2,188 (47.6)	
2	6,013 (32.7)	1,347 (29.3)	3,062 (33.3)	1,586 (34.5)	
Age at diagnosis, yr	1.9 (3.3)	1.6 (3.0)	1.9 (3.3)	2.1 (3.4)	<0.0001
Age at baseline, yr	9.8 (4.1)	12.0 (4.6)	9.5 (3.8)	8.6 (3.2)	<0.0001
Birth cohort					<0.0001
<1981	2,427 (13.2)	1,315 (28.6)	965 (10.5)	244 (5.3)	
1981–1988	6,454 (35.1)	1,958 (42.6)	3,255 (35.4)	1,282 (27.9)	
1989–1994	4,505 (24.5)	772 (16.8)	2,317 (25.2)	1,365 (29.7)	
>1994	5,001 (27.2)	556 (12.1)	2,657 (28.9)	1,705 (37.1)	
FEV_1, % predicted	86.1 (21.0)	62.7 (17.8)	87.5 (13.9)	105.2 (12.9)	<0.0001
BMI, percentile	45.0 (26.8)	30.3 (24.6)	45.3 (25.8)	57.2 (24.3)	<0.0001
Pancreatic enzyme use	17,137 (93.2)	4,417 (96.1)	8,569 (93.2)	4,159 (90.5)	<0.0001
Acute exacerbations					<0.0001
0	13,809 (75.1)	2,335 (50.9)	7,318 (79.6)	4,022 (87.5)	
1	2,206 (12.0)	812 (17.7)	1,067 (11.6)	363 (7.9)	
2	1,342 (7.3)	674 (14.7)	552 (6.0)	156 (3.4)	
⩾3	1,030 (5.6)	766 (16.7)	257 (2.8)	55 (1.2)	
Infections					
MRSA					
At baseline	846 (4.6)	280 (6.1)	414 (4.5)	165 (3.6)	<0.0001
Never infected during follow-up	12,099 (65.8)	3,043 (66.2)	5,912 (64.3)	3,139 (68.3)	<0.0001
Pseudomonas aeruginosa					
At baseline	8,219 (44.7)	3,163 (68.8)	3,871 (42.1)	1,319 (28.7)	<0.0001
Never infected during follow-up	2,887 (15.7)	301 (6.55)	1,425 (15.5)	1,117 (24.3)	<0.0001
Burkholderia cepacia					
At baseline	294 (1.6)	188 (4.1)	101 (1.1)	18 (0.4)	<0.0001
Never infected during follow-up	17,026 (92.6)	4,059 (88.3)	8,560 (93.1)	4,380 (95.3)	<0.0001
ABPA					
At baseline	294 (1.6)	143 (3.1)	129 (1.4)	41 (0.9)	<0.0001
Not acquired during follow-up	16,493 (89.7)	3,875 (84.3)	8,275 (90.0)	4,302 (93.6)	<0.0001
NTM					
At baseline	92 (0.5)	46 (1.0)	37 (0.4)	9 (0.2)	<0.0001
Not present during follow-up	17,247 (93.8)	4,243 (92.3)	8,596 (93.5)	4,394 (95.6)	<0.0001
Stenotrophomonas					
At baseline	1,030 (5.6)	336 (7.3)	515 (5.6)	184 (4.0)	<0.0001
Never infected during follow-up	11,326 (61.6)	2,846 (61.9)	5,489 (59.7)	2,992 (65.1)	<0.0001
CFRD					
At baseline	37 (0.2)	14 (0.3)	18 (0.2)	9 (0.2)	0.7
Ever during follow-up	2,850 (15.5)	1,048 (22.8)	1,388 (15.1)	455 (9.9)	<0.0001
Lower SES					
At baseline	7,980 (43.4)	2,409 (52.4)	3,972 (43.2)	1,655 (36.0)	<0.0001
Ever during follow-up	11,565 (62.9)	3,342 (72.7)	5,737 (62.4)	2,551 (55.5)	<0.0001
Alive	17,615 (95.8)	3,861 (84.0)	9,111 (99.1)	4,582 (99.7)	<0.0001

Definition of abbreviations: ABPA = allergic bronchopulmonary aspergillosis; BMI = body mass index; CFRD = cystic fibrosis–related diabetes; MRSA = methicillin-resistant *Staphylococcus aureus*; NTM = nontuberculous mycobacteria; SES = socioeconomic status.
Continuous variables are summarized as mean (SD); categorical variables are summarized as n (%). *P* values, obtained from chi-square tests for categorical data and linear models for continuous data, reflect overall associations.
*Phenotype corresponds to late, middle, and early rapid lung function decline.

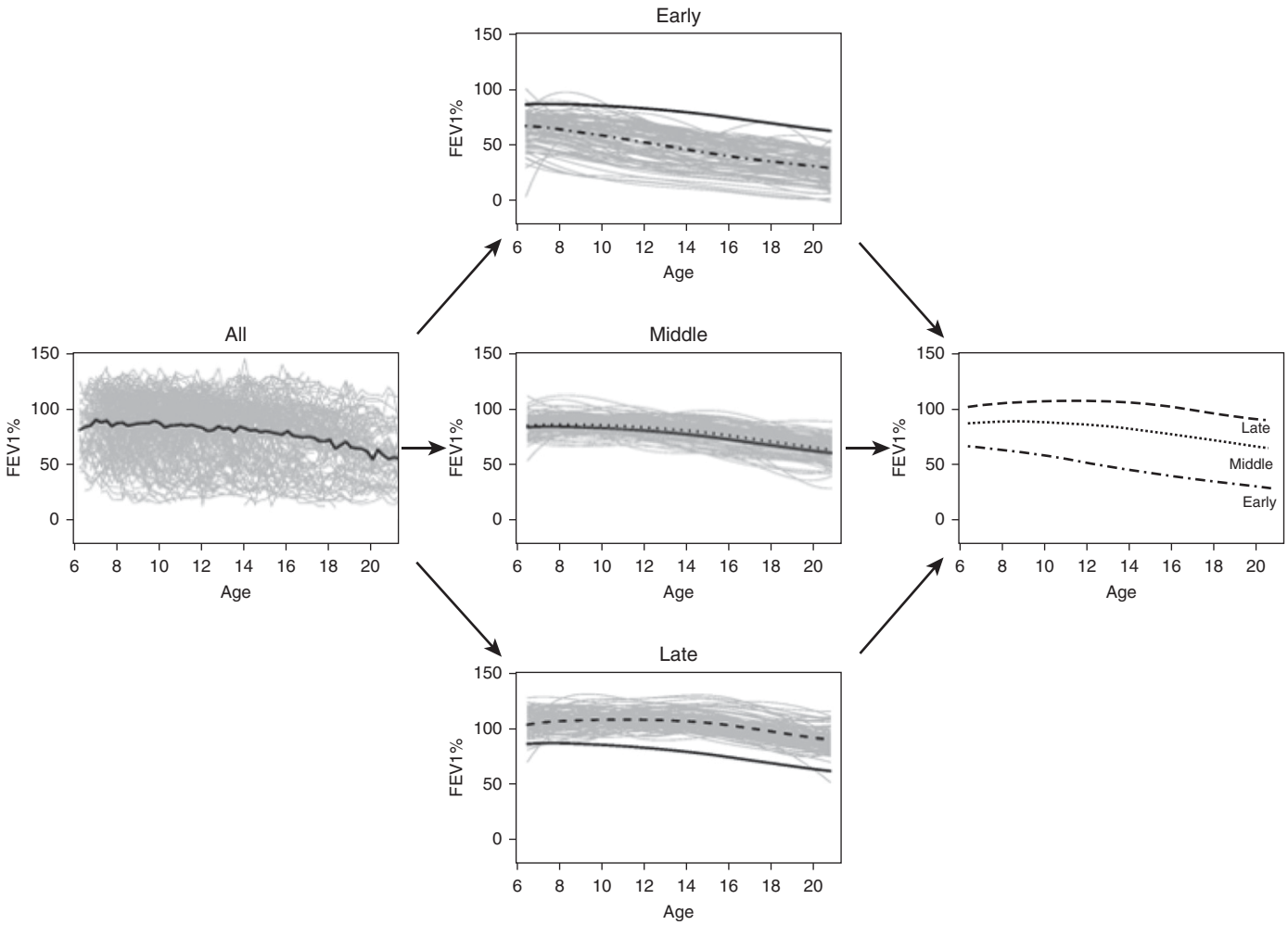

Figure 1. Trajectories of FEV_1 decline (expressed as % predicted) in patients with cystic fibrosis aged 6–21 years (*left plot*). Phenotypes of rapid decline are segmented by functional principal components analysis with the *solid black line* as a reference to the population-level average decline in FEV_1 over age (*middle plots*); the average FEV_1 progression for each phenotype is shown (*right plot*).

bronchopulmonary aspergillosis (ABPA), CF-related diabetes (CFRD), and lower socioeconomic status as indicated by use of Medicaid insurance. For each infection/microbiology variable, presence was indicated as "1," whereas "0" or missing entries were assumed an absence of infection.

Evaluating associations between longitudinal markers and phenotypes. To examine temporal associations between other time-varying characteristics and phenotype membership during follow-up, we developed longitudinal models using generalized estimating equations, each with an appropriate link function for the categorical or continuous outcome of interest. In each model, phenotype membership was included as a covariate. Longitudinal outcomes were BMI percentile, detection of MRSA, *P. aeruginosa*, *B. cepacia*, ABPA, NTM, *Stenotrophomonas*, and diagnosis of CFRD. Longitudinal entries with missing covariate information were excluded from regression models.

Results

Cohort Characteristics

The analysis cohort data was comprised of 565,401 FEV_1 observations on 18,387 patients with data from 1997 to 2013 in the CFFPR (Table 1). The median (range) number of FEV_1 observations per individual patient was 19 (7–55), with a follow-up period of 6.8 (1.5–15.5) years and a total of 178,635 person-years of follow-up. There were 2,624 patients with less than seven FEV_1 observations excluded from the primary analyses. Compared with their counterparts in the analysis cohort, these patients were slightly older at baseline, on average, but typically belonged to the youngest birth cohort (*see* Table E3 in the online supplement). Both cohorts had similar mean FEV_1 and socioeconomic status levels at entry, but the excluded cohort had slightly lower prevalence of infections and higher BMI percentile.

Longitudinal FEV_1 Pattern Classification

The first FPC, which characterizes how each individual patient's lung function trajectory differs from the mean trajectory, explained 94% of the variation in FEV_1. Each of the three clusters, created using scores from the first FPC (the distribution is shown in Figure E1), corresponded to a late, middle, or early phenotype of rapid decline. Based on descriptive analysis, patients with the late decline phenotype were, on average, diagnosed later, had better nutrition status, fewer acute exacerbations, less infections,

and higher socioeconomic status than their middle and early decline counterparts (Table 1). Classifications that further refined phenotypes based on the first FPC and incorporated the second FPC are reported as supplemental material (Section E2). Including the second FPC, which characterizes higher-order modes of variation in FEV_1 trajectories, was able to explain roughly 97% of the variation in FEV_1. Although statistically significant, bivariate associations between duration of follow-up and FPC scores had relatively small correlation coefficients (Pearson r, −0.17 and 0.04 for FPC scores from the first and second components, respectively).

The FEV_1 data exhibited substantial variation among patients and also within each patient over time (Figure 1, *left*). Patient-specific trajectories obtained from sparse longitudinal FPCA, regardless of phenotype membership, demonstrated progressive loss of lung function over age. The late and middle phenotypes of lung function decline (Figure 1, *middle*) seemed to have similar between-patient variation in FEV_1 over age, but both phenotypes exhibited less variation than the early decline phenotype. Mean FEV_1 trajectories for the late and middle phenotypes had similar shapes, but the middle phenotype exhibited lung function that was consistently lower. Of all the phenotypes, patients with earliest decline had the lowest FEV_1 initially, with loss that tended to progress more rapidly over time. These phenotypes were further classified by examining quintiles of the first FPC (*see* Figure E2) and using the second FPC (*see* Figure E3).

Degree and timing of rapid decline differed according to phenotype, with patients who had the early decline phenotype continually losing lung function (the curve estimating the mean rate of change was always below zero); average maximal loss was 3.2% predicted/yr and occurred at age 12.9 years (Figure 2, *left*). Patients with the middle decline phenotype experienced an average maximal FEV_1 loss of 2.8% predicted/yr but at an older age (16.3 yr) (Figure 2, *middle*). Those who had the late decline phenotype experienced an average maximal FEV_1 loss of 2.9% predicted/yr at 18.5 years of age (Figure 2, *right*). Although degrees of maximal loss among phenotypes were similar, the group exhibiting rapid decline later in life experienced the greatest overall loss of lung function. Age and degree of maximal loss were more variable across classifications based on quintiles of the first FPC (*see* Figure E4), and those that included the second FPC (*see* Figure E5); timing of most rapid FEV_1 decline ranged from 7.8 to 19.2 years, and maximal FEV_1 loss ranged from 2.7% to 5.7% predicted/yr (*see* Table E2).

Performing FPCA on a larger cohort that included data from patients with at least two measurements yielded similar results in terms of the distribution of FPC scores and proportion of explainable variance per component (*see* Section E4). Although findings were similar, the model required more complex curve-fitting parameters. This is likely caused by the additional uncertainty introduced by including profiles with fewer FEV_1 observations.

Predictors of Early Phenotype of Lung Function Decline

Individuals at risk of earliest rapid lung function decline tended to be females from older birth cohorts who had lower FEV_1, BMI percentile, and more frequent acute exacerbations at entry (Table 2). Of all infections, *B. cepacia* at entry corresponded to the highest odds of having an early decline phenotype, followed by MRSA and *P. aeruginosa*. ABPA and NTM infections were not statistically significant predictors of early phenotype membership. Although *Stenotrophomonas* infection was associated with higher odds of developing the early decline phenotype, CFRD status and pancreatic enzyme use were not significant predictors in the model.

Associations between Longitudinal Markers and Phenotypes of Lung Function Decline

Individuals with the late decline phenotype had the highest BMI percentile, on average, followed by their middle and early decline counterparts (Table 3). Prevalence of infections with *P aeruginosa* and *B. cepacia*, and diagnosis of CFRD, tended to have a dose–response association with phenotypes, increasing across late, middle, and early phenotypes. MRSA infection prevalence was significantly higher in individuals with the early decline phenotype, compared with those in the late decline phenotype.

Discussion

We have shown that a novel method well suited for clustering longitudinal lung

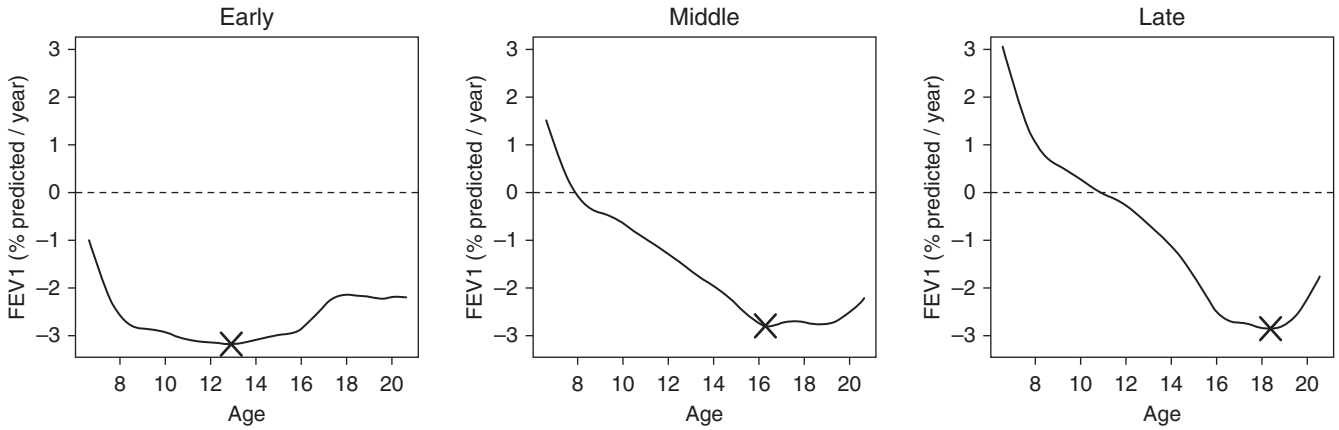

Figure 2. Trajectories of rate of lung function decline (expressed as % predicted/yr) over age classified according to early, middle, and late phenotypes of rapid lung function decline. *X* marks the point of maximal FEV_1 loss over age for each phenotype. Portions of the curve *above* and *below* the *horizontal dashed line* at zero imply that FEV_1 is increasing and decreasing over age, respectively.

Table 2. Baseline Characteristics as Predictors of Early Rapid Decline Phenotype

Variable	OR*	95% CI		*P* Value
Age at diagnosis	0.945	0.927	0.963	<0.0001
Age at baseline	0.957	0.930	0.984	0.0023
Δ F508 copies				
None	1.733	1.358	2.212	<0.0001
1	0.991	0.802	1.225	0.0016
2	1			
Birth cohort				
<1981	3.962	2.789	5.629	<0.0001
1981–1988	3.940	3.177	4.886	<0.0001
1989–1994	1.606	1.318	1.957	<0.0001
>1994	1			
Male sex	0.773	0.693	0.863	<0.0001
$FEV_{1\%}$ predicted	0.909	0.905	0.913	<0.0001
BMI percentile	0.993	0.990	0.995	<0.0001
Pancreatic enzyme use	0.982	0.743	1.297	0.8
Acute exacerbations				
0	0.432	0.348	0.536	<0.0001
1	0.607	0.476	0.775	0.2
2	0.717	0.554	0.929	0.2
⩾3	1			
Infections				
MRSA	1.691	1.308	2.186	<0.0001
Pseudomonas aeruginosa	1.658	1.476	1.861	<0.0001
Burkholderia cepacia	2.067	1.431	2.985	0.0001
ABPA	1.360	0.950	1.948	0.07
NTM	1.392	0.715	2.711	0.3
Stenotrophomonas	1.282	1.028	1.599	0.0209
CFRD	0.881	0.270	2.877	0.8
Lower SES	1.198	1.073	1.337	0.0027

Definition of abbreviations: ABPA = allergic bronchopulmonary aspergillosis; BMI = body mass index; CFRD = cystic fibrosis–related diabetes; CI = confidence interval; MRSA = methicillin-resistant *Staphylococcus aureus*; NTM = nontuberculous mycobacteria; OR = odds ratio; SES = socioeconomic status.

*For categorical variables, the estimate is the OR, expressed as the odds of having the rapid decline phenotype for the indicated group versus the odds for the reference group (labeled by estimate = 0). For continuous covariates, the OR is the change in odds of having the rapid decline phenotype when the covariate increases by one unit. The last columns provide the 95% CI for each variable and corresponding *P* value. An OR >1 indicates association with a higher risk of having the rapid decline phenotype.

function measurements from a national patient registry provides new insights into the understanding of CF lung disease progression. Advanced statistical models have aided in predicting the natural history of CF lung disease, bringing the field largely to the consensus that rapid decline often occurs sometime between adolescence and early adulthood; however, the specific timing of rapid decline varies across studies. For example, Liou and colleagues (16) report maximal decline in 14–15 year olds, whereas Vandenbranden and colleagues (5) found it to be during early adulthood. Although discrepancies among these and other studies could be caused in part by differences in analytic approaches and datasets, results of these studies demonstrate the existence of substantial heterogeneity among the FEV_1 trajectories. Our study illuminates phenotypes inherent in these findings and characterizes their respective clinical courses in CF. Results of this study have implications for both CF clinical care and research.

By understanding similarities and differences among patterns of longitudinal FEV_1, our approach provides opportunities for more timely and targeted intervention for patients and phenotypes at risk of rapid lung function decline. With the increased emphasis on development of biomarkers to detect early stage CF lung disease, having defined phenotypes of lung function decline may improve the understanding of associations between novel markers, such as imaging measures of lung structure and function (17), and existing surrogates, such as FEV_1. Phenotypes of rapid lung function decline may be used to identify individuals with the greatest potential to benefit from new therapeutics or to optimize use of available interventions. For example, individuals in the most severe phenotype who experience early and sustained loss of lung function from adolescence until early adulthood may be better candidates for the early initiation of specific interventions compared with their late or middle decline onset counterparts. In addition, proactive indication of characteristics associated with early decline, such as acute pulmonary exacerbations, sex, BMI, and microbiology, may inform caregivers of "highest risk" patients who may require unique approaches and monitoring to maintain lung function and overall health; actions could include increased visit frequency and implementation of nocturnal feeding.

Although these additional interventions are not currently part of CF care guidelines for rapid decline, they represent opportunities for discussions with families and patients about optimizing their own care. Future studies linking the rapid decline phenotype to predictive molecular biomarkers may allow caregivers and patients to assess the relative benefit of different interventions specific to their own care plan. Cohort selection according to mean lung function or slope may lead to classifications that do not identify those at risk of maximal loss of lung function, or the timing at which decline becomes established. This becomes increasingly important in an era of personalized care in which nonselective application may not provide uniform benefit.

Results of this study further substantiate a growing body of literature that suggests lung function decline in CF is nonlinear (18). This is evident in Figure 2, where loss of lung function varies over time within each group and between groups. Konstan and colleagues (4), among others, have helped us understand contributors to lung function decline in CF, such as being male or having *P. aeruginosa* infection. Our evaluation of rapid FEV_1 decline builds on the most recent linear change-point estimates reported by Moss and colleagues (7). They identified a "relatively stable" cluster of individuals with FEV_1 loss of 0.5% predicted/yr and a majority of individuals with a loss 4.4% predicted/yr starting at 14.6 years of age. Although theirs was a single-center study assessing change points to estimate rate of decline, our FPCA results assuming a smooth, continuous rate of decline yielded a portion of clusters with individuals exhibiting similar patterns of decline to those identified by Moss and coworkers (7) (*see* Figure E5). Furthermore,

Table 3. Temporal Associations with Phenotypes of Lung Function Decline

Longitudinal Outcome*	Estimate (95% CI) by Phenotype		
	Early	Middle	Late
BMI, percentile	25.2 (24.6–25.8)	41.8 (40.4–43.2)	55.2 (53.7–56.7)
Infections, prevalence			
MRSA	17.2 (16.0–18.2)	15.9 (13.8–18.3)	12.6 (10.8–14.7)
Pseudomonas aeruginosa	71.2 (69.8–71.9)	52.1 (49.6–50.0)	39.8 (38.1–41.9)
Burkholderia cepacia	5.7 (5.1–6.4)	2.2 (1.68–2.9)	1.2 (0.9–1.6)
ABPA	6.3 (5.7–6.9)	3.8 (3.1–4.8)	2.5 (1.9–3.3)
NTM	1.9 (1.6–2.2)	1.2 (0.9–1.7)	0.8 (0.6–1.1)
Stenotrophomonas	13.9 (13.0–14.7)	12.7 (11.1–14.5)	9.6 (8.3–11.1)
CFRD, prevalence	9.8 (9.2–10.5)	6.2 (5.3–7.3)	4.0 (3.3–4.8)

Definition of abbreviations: ABPA = allergic bronchopulmonary aspergillosis; BMI = body mass index; CFRD = cystic fibrosis–related diabetes; CI = confidence interval; MRSA = methicillin-resistant *Staphylococcus aureus*; NTM = nontuberculous mycobacteria.
*Estimates (95% CI) in each row are from separate models with phenotype membership as a covariate and adjustment for longitudinal correlation (*see* METHODS). For the first row, the estimates are the mean (95% CI) percentile; for subsequent rows, each estimate corresponds to prevalence.

this study confirms the "ceiling effect" previously found by Konstan and colleagues (4) in analysis of another CF registry; children and adolescents with the highest initial FEV_1 had demonstrated the steepest decline.

Phenotypes of rapid decline onset were identified from this study using a functional data analysis method, as opposed to traditional cluster analysis. We considered the temporal information from the FEV_1 functional data, and have classified the fitted FEV_1 curves according to early, middle, and late onset of rapid decline. As shown in the online supplement, the scores provided by this analysis can be used to examine a range of phenotypes further segmented than the three groups shown. FPCA has also been implemented in medical monitoring studies to characterize variation in longitudinal trajectories; applications range from characterizing glycemic control (13) to electroencephalographic activity during sleep (19). Clustering of lung function data has also been used to identify clinically relevant asthma phenotypes (20–22).

Our study has some limitations that should be considered, including potential survivor bias (Table 1; *see* Figure E4). Given the age range of the cohort studied (6–21 yr), the impact of informative dropout should be lessened, compared with studies across the full age range that would be subject to increased drop out because of death. Our approach assumes that data are missing at random (23). A broader age range could be included to assess patient trajectories according to later-stage lung function decline, but would need to account for survivor effects shown in previous work (24). There is a clear birth cohort effect on the model results, which has been noted in other CF registry studies (2, 6, 25) and could be reflective of advancements in care that were largely unavailable to older patients with the early decline phenotype. To fully characterize rapid decline in the most modern era of CF care would require prospective data on the youngest individuals in the cohort. Future studies of rapid decline phenotypes should incorporate race and ethnicity, considering recent findings that Hispanic patients with CF have higher mortality rates than their non-Hispanic counterparts (26). Although not studied here, phenotypes could be used to further assess treatment selection bias found in previous clinical effectiveness studies of mortality and pulmonary decline. Examples include longitudinal studies of tobramycin effectiveness using the CF registry (27–29). A portion of the results could be attributable to incomplete data, such as the lack of association between ABPA or NTM infection and early phenotype. FEV_1 trajectories may differ according to the type of spirometry reference equation applied to the raw FEV_1 data (30). Using CFFPR data for patients aged 8–17 years taken from 2013, Wang and Hankinson equations yielded higher median $FEV_{1\%}$ predicted values, compared with values obtained using the Global Lung Initiative Equations (31). Monitoring changes in FEV_1, however, seem to be less susceptible to this effect.

In conclusion, we have used novel statistical modeling on a national registry to identify phenotypes of CF adolescents at risk for early, middle, and late onset of rapid lung function decline. Key predictors of the early decline phenotype included female sex, airway microbiology (MRSA, *P. aeruginosa*, and *B. cepacia*), a positive history of pulmonary exacerbation and birth cohort. Although the rate of maximal decline was similar across early, middle, and late phenotypes, lung function trajectories were not. Phenotypes identified in this analysis could be useful in prognostic care of individuals with CF and research studies aimed at targeting individuals who may maximally benefit from particular interventions at a certain point in their disease progression. Understanding the phenotypes of patients with CF with early, middle, and late onset of rapid decline may further pave the way to improved predictive algorithms for CF clinical care and research. ■

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

Acknowledgment: The authors thank the Cystic Fibrosis Foundation Patient Registry Committee for their thoughtful comments and data dispensation and Dr. Patrick Ryan for his critical review of the manuscript.

References

1. Liou TG, Adler FR, Cox DR, Cahill BC. Lung transplantation and survival in children with cystic fibrosis. *N Engl J Med* 2008;359:2143–2152.
2. Taylor-Robinson D, Whitehead M, Diderichsen F, Olesen HV, Pressler T, Smyth RL, Diggle P. Understanding the natural progression in %FEV1 decline in patients with cystic fibrosis: a longitudinal study. *Thorax* 2012;67:860–866.
3. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, Coates A, van der Grinten CP, Gustafsson P, Hankinson J, *et al*. Interpretative strategies for lung function tests. *Eur Respir J* 2005;26: 948–968.
4. Konstan MW, Morgan WJ, Butler SM, Pasta DJ, Craib ML, Silva SJ, Stokes DC, Wohl ME, Wagener JS, Regelmann WE, *et al*.; Scientific Advisory Group and the Investigators and Coordinators of the Epidemiologic Study of Cystic Fibrosis. Risk factors for rate of decline

in forced expiratory volume in one second in children and adolescents with cystic fibrosis. *J Pediatr* 2007;151:134–139, 139.e1.
5. Vandenbranden SL, McMullen A, Schechter MS, Pasta DJ, Michaelis RL, Konstan MW, Wagener JS, Morgan WJ, McColley SA; Investigators and Coordinators of the Epidemiologic Study of Cystic Fibrosis. Lung function decline from adolescence to young adulthood in cystic fibrosis. *Pediatr Pulmonol* 2012;47:135–143.
6. Szczesniak RD, McPhail GL, Duan LL, Macaluso M, Amin RS, Clancy JP. A semiparametric approach to estimate rapid lung function decline in cystic fibrosis. *Ann Epidemiol* 2013;23:771–777.
7. Moss A, Juarez-Colunga E, Nathoo F, Wagner B, Sagel S. A comparison of change point models with application to longitudinal lung function measurements in children with cystic fibrosis. *Stat Med* 2016;35: 2058–2073.
8. James GM, Hastie TJ, Sugar CA. Principal component models for sparse functional data. *Biometrika* 2000;87:587–602.
9. Szczesniak RD, Li D, Clancy JP. Finding FEV1 phenotypes of rapid decline during adolescence and young adulthood. Poster Session Abstracts, No. 497. *Pediatr Pulmonol* 2016;51(S45):S194–S485.
10. Knapp EA, Fink AK, Goss CH, Sewall A, Ostrenga J, Dowd C, Elbert A, Petren KM, Marshall BC. The Cystic Fibrosis Foundation Patient Registry: design and methods of a national observational disease registry. *Ann Am Thorac Soc* 2016;13:1173–1179.
11. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;159:179–187.
12. Wang X, Dockery DW, Wypij D, Fay ME, Ferris BG Jr. Pulmonary function between 6 and 18 years of age. *Pediatr Pulmonol* 1993;15: 75–88.
13. Szczesniak RD, Li D, Duan LL, Altaye M, Miodovnik M, Khoury JC. Longitudinal patterns of glycemic control and blood pressure in pregnant women with type 1 diabetes mellitus: phenotypes from functional data analysis. *Am J Perinatol* 2016;33:1282–1290.
14. Kim J-o, Mueller CW. Factor analysis: statistical methods and practical issues. Beverly Hills, CA: Sage Publications; 1978. p. 88.
15. Frøslie KF, Røislien J, Qvigstad E, Godang K, Bollerslev J, Voldner N, Henriksen T, Veierød MB. Shape information from glucose curves: functional data analysis compared with traditional summary measures. *BMC Med Res Methodol* 2013;13:6.
16. Liou TG, Elkin EP, Pasta DJ, Jacobs JR, Konstan MW, Morgan WJ, Wagener JS. Year-to-year changes in lung function in individuals with cystic fibrosis. *J Cyst Fibros* 2010;9:250–256.
17. Fleck R, McPhail G, Szczesniak R, Knowlton J, Radhakrishnan R, Clancy J, Amin R. Aortopulmonary collateral flow in cystic fibrosis assessed with phase-contrast MRI. *Pediatr Radiol* 2013;43: 1279–1286.
18. Harun SN, Wainwright C, Klein K, Hennig S. A systematic review of studies examining the rate of lung function decline in patients with cystic fibrosis. *Paediatr Respir Rev* 2016;20:55–66.
19. Crainiceanu CM, Caffo BS, Di CZ, Punjabi NM. Nonparametric signal extraction and measurement error in the analysis of electroencephalographic activity during sleep. *J Am Stat Assoc* 2009;104:541–555.
20. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, Wardlaw AJ, Green RH. Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med* 2008;178:218–224.
21. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, D'Agostino R Jr, Castro M, Curran-Everett D, Fitzpatrick AM, *et al*.; National Heart, Lung, and Blood Institute's Severe Asthma Research Program. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med* 2010; 181:315–323.
22. Bafadhel M, McKenna S, Terry S, Mistry V, Reid C, Haldar P, McCormick M, Haldar K, Kebadze T, Duvoix A, *et al*. Acute exacerbations of chronic obstructive pulmonary disease: identification of biologic clusters and their biomarkers. *Am J Respir Crit Care Med* 2011;184:662–671.
23. Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, editors. Longitudinal data analysis. Boca Raton, FL: Taylor & Francis Group, LLC; 2009.
24. Szczesniak RD, McPhail GL, Li D, Amin RS, Clancy JP. Predicting future lung function decline in cystic fibrosis patients: Statistical methods and clinical connections. *Pediatr Pulmonol* 2016;51:217–218.
25. VanDevanter DR, Pasta DJ, Konstan MW. Improvements in lung function and height among cohorts of 6-year-olds with cystic fibrosis from 1994 to 2012. *J Pediatr* 2014;165:1091–1097.e2.
26. Buu MC, Sanders LM, Mayo JA, Milla CE, Wise PH. Assessing differences in mortality rates and risk factors between Hispanic and non-Hispanic patients with cystic fibrosis in California. *Chest* 2016; 149:380–389.
27. Rothman KJ, Wentworth CE III. Mortality of cystic fibrosis patients treated with tobramycin solution for inhalation. *Epidemiology* 2003; 14:55–59.
28. Sawicki GS, Signorovitch JE, Zhang J, Latremouille-Viau D, von Wartburg M, Wu EQ, Shi L. Reduced mortality in cystic fibrosis patients treated with tobramycin inhalation solution. *Pediatr Pulmonol* 2012;47:44–52.
29. VanDyke RD, McPhail GL, Huang B, Fenchel MC, Amin RS, Carle AC, Chini BA, Seid M. Inhaled tobramycin effectively reduces FEV_1 decline in cystic fibrosis: an instrumental variables analysis. *Ann Am Thorac Soc* 2013;10:205–212.
30. Stanojevic S, Stocks J, Bountziouka V, Aurora P, Kirkby J, Bourke S, Carr SB, Gunn E, Prasad A, Rosenfeld M, *et al*. The impact of switching to the new global lung function initiative equations on spirometry results in the UK CF registry. *J Cyst Fibros* 2014;13: 319–327.
31. Cystic Fibrosis Foundation. Cystic Fibrosis Foundation Patient Registry annual report. Bethesda, MD: Cystic Fibrosis Foundation; 2014.