Classifying Severity of Cystic Fibrosis Lung Disease Using Longitudinal Pulmonary Function Data

Mark D. Schluchter, Michael W. Konstan, Mitchell L. Drumm, James R. Yankaskas, and Michael R. Knowles

Departments of Pediatrics and Genetics, Case Western Reserve University School of Medicine, Cleveland, Ohio; and Cystic Fibrosis–Pulmonary Research and Treatment Center, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Rationale: The study of genetic modifiers in cystic fibrosis (CF) lung disease requires rigorous phenotyping. One type of genetic association study design compares polymorphisms in patients at extremes of phenotype, requiring accurate classification of pulmonary disease at varying ages.

Objective: To evaluate approaches to quantify severity of pulmonary disease and their ability to discriminate between patients with CF at the extremes of phenotype.

Methods: $\Delta F508$ homozygotes ($n = 828$) were initially classified as “severe” (approximate lowest quartile of FEV_1 (% pred) for age, 8–25 yr) or “mild” disease (highest quartile of FEV_1 for age, ≥ 15 yr). FEV_1 measurements from the 5 yr before enrollment (total = 18,501 measurements; average 23 per subject) were analyzed with mixed models, and patient-specific estimates of FEV_1 (% pred) at ages 5, 10, 15, 20, and 25 yr and slope of FEV_1 versus age were examined for their ability to discriminate between groups using receiver operating characteristics (ROC) curve areas.

Results: Logistic regression of severity group on mixed model (empirical Bayes) estimates of intercept and slope of FEV_1 (% pred) versus age discriminated better than did classification using FEV_1 slope alone (ROC area = 0.995 vs. 0.821) and was equivalent to using estimated FEV_1 at 20 yr of age as a single discriminator. The estimated survival percentile from a joint survival/longitudinal model provided equally good classification (ROC area = 0.994).

Conclusions: In CF, estimated FEV_1 (% pred) at 20 yr of age and the estimated survival percentile are useful indices of pulmonary disease severity.

Keywords: association studies; FEV_1 ; genetic modifiers

Cystic fibrosis (CF) is an autosomal recessive genetic disease caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. Because most patients with CF develop progressive pulmonary disease, measures of pulmonary involvement, in particular FEV_1 , have been used as markers of disease severity and to predict survival (1–3). However, considerable heterogeneity exists in prognosis and severity, even among patients of the same genotype, suggesting that genetic modifiers may play a role (4). The study of gene modifiers in monogenic disorders requires rigorous phenotyping, and inadequate characterization of the phenotype is a well-recognized limitation of case-control genetic association studies (5–8). In CF, pulmonary phenotyping is complex because the progression of lung disease

is likely multifactorial and is multiphasic or nonlinear for some patients (9, 10).

One specific type of case-control genetic association study design compares patients at the extremes of phenotype (i.e., those with severe versus mild disease) because it provides additional power to detect gene modifiers (11). In such studies, it is particularly important to be able to accurately identify patients at the extremes of phenotype. In this article, we investigate approaches to classifying severity of disease when longitudinal lung function measures are available and compare them using data from the Gene Modifier Study (GMS), a large, multicenter study of genetic modifiers of CF lung disease (12). The goal of the GSM is to examine the association of genetic polymorphisms with pulmonary phenotype in $\Delta F508$ homozygotes by comparing patients with “severe” versus “mild” pulmonary disease.

METHODS

Design and Enrollment Criteria in the GSM

Patients were initially enrolled into “severe” or “mild” groups based on current age and most recent FEV_1 (% pred) while they were clinically stable (12). The severe group included patients 8–25 yr of age, and the mild group was divided into young and older subgroups (15–28 and ≥ 29 yr of age, respectively). Enrollment by severity group was based on age-specific cutoff values for FEV_1 (% pred; see online supplement), derived approximately from quartiles of lung function for patients < 34 yr of age in the U.S. Cystic Fibrosis Foundation Patient Registry (13). Patients ≥ 34 yr of age were considered “mild” regardless of their FEV_1 , on the basis of survival. All available spirometric data in the previous 5 yr for each patient were collected, including pre- and post-bronchodilator values, regardless of the patient’s status with respect to acute illnesses. Analyses used the prebronchodilator value if available (97% of measurements). Otherwise, the post-bronchodilator value was used. Percent predicted values for spirometry measurements were calculated using Knudson equations (14).

Characterization of Longitudinal Patterns of FEV_1 in Individual Patients

Longitudinal FEV_1 data were analyzed using a mixed model, assuming that the mean FEV_1 (% pred) follows a linear regression versus time for each patient with random patient-specific slope and intercept, with a separate population regression line for each of the three severity groups (Appendix 1). The mixed model provided estimates of the mean intercept and slope by severity group and estimates of patient-specific slopes and intercepts (empirical Bayes estimates).

We first reexamined the initial classification of each patient, which was based on their single FEV_1 at enrollment, by analyzing all longitudinal data. Mixed model estimates of individual patients’ intercepts (FEV_1 at age of birth) and slopes (rates of FEV_1 decline), obtained from the mixed linear model fit to data from all patients, were used to estimate the true underlying FEV_1 at enrollment for each subject. A logistic regression of disease group (severe vs. mild) on age and estimated FEV_1 at enrollment was fit, and patients for whom the predicted probability of being in their initial group was less than 0.90 were identified as outliers and excluded from further analyses. This was done using all FEV_1 values and by using the maximal yearly value for each patient. This resulted in a final dataset with 802 subjects (256 severe, 304 young mild, 242 older mild) with a total of 18,501 FEV_1 measurements.

(Received in original form December 16, 2005; accepted in final form July 14, 2006)

Supported by Cystic Fibrosis Foundation (CFF) grant KNOWLE00A0 (M.R.K.), CFF grant DRUMM00A0 (M.L.D.), and NIH grant HL68890 (M.R.K.).

Correspondence and requests for reprints should be addressed to Mark D. Schluchter, Ph.D., Department of Epidemiology and Biostatistics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106–4945. E-mail: mds11@case.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 174, pp 780–786, 2006

Originally Published in Press as DOI: 10.1164/rccm.200512-1919OC on July 20, 2006
Internet address: www.atsjournals.org

Figure 1. FEV₁ criteria (boundaries) used to enroll patients as severe (8–25 yr of age) or mild (≥ 15 yr of age) in the Gene Modifier Study. Specific cut-off values are provided in the online supplement. The dotted line shows median FEV₁ (% pred) for age for patients with cystic fibrosis (CF) who were 12–24 yr of age (average of male and female values), using CF-specific reference standards (22).

Receiver operating characteristics (ROC) curves (15) were used to evaluate and compare several patient-specific summary indices to assess which were best at discriminating between mild and severe groups. The summary indices included mixed model and least squares (LS) estimates of patient-specific rates of decline in FEV₁ (% pred) and levels of FEV₁ (% pred) at 5–25 yr of age.

We predicted age at death based on a model relating survival to longitudinal lung function, using parameters estimated from an analysis of an external dataset of $n = 188$ homozygous $\Delta F508$ patients (2). This estimate of age at death was also evaluated as a severity measure.

Analyses were performed using SAS version 9.1 (16). Data are presented as mean $\pm$ SD. Between-group comparisons were made using χ^2 tests for categorical variables and using one-way analysis of variance for continuous variables.

RESULTS

Screening for Outliers and Characterization of Severity Groups

The age-specific boundaries of FEV₁ (% pred) at enrollment, used for initial classification by severity, are shown in Figure 1. The specific cutoffs used at each age are provided in the online supplement.

Mixed model estimates of slope and intercept (at birth; i.e., age, 0 yr) of the regression of FEV₁ (% pred) versus age were used to estimate the level of FEV₁ at the age of enrollment for each patient and at other fixed ages (e.g., 15, 10, 15, 20, and 25 yr). In Figure 2, the estimated FEV₁ (% pred) values at enrollment (obtained from an analysis of all FEV₁ values) are plotted against age of enrollment for the 820 patients. Patients identified as outliers based on logistic regression of severity group on age and estimated FEV₁ (% pred) at enrollment (see METHODS) are identified as open circles ($n = 12$). Similar analysis using only the best yearly FEV₁ values identified an additional six patients as outliers, labeled as open triangles in the figure. These 18 patients were excluded from further analyses reported herein.

Table 1 summarizes demographic and clinical information at the time of enrollment and the number of FEV₁ measurements obtained per subject in the 5 yr before enrollment for the 802 patients separated into severe, young mild, and older mild groups. Summary indices describing the individual patients' regression lines of FEV₁ (% pred) versus age are also provided, including the patient-specific slope and estimated FEV₁ (% pred) at 15 and 20 yr of age. The difference between mean predicted

levels of FEV₁ (% pred) at 15 and 20 yr of age within a group equals the mean slope multiplied by 5 yr. Significant differences between all three groups were found in terms of FEV₁ decline and estimated levels of FEV₁ at 15 and 20 yr of age, with the severe and young mild groups being most and least severe, respectively, and the older mild group having values intermediate between the other two groups.

Ability of Measures of Disease Severity to Discriminate between Extremes of Phenotype

Patient-specific summary measures, derived from the patients' longitudinal data, were evaluated in their ability to distinguish between severe versus mild groups. Discriminating ability was ascertained using the area under the ROC curve, which plots sensitivity versus 1-specificity of the index as the threshold value (cutpoint) used for classification is varied. The ROC area ranges

Figure 2. Estimated FEV₁ (% pred) at age at enrollment versus age at enrollment. The FEV₁ (% pred) values were estimated based on a mixed model analysis of all available data from each patient (see text and Appendix 1). Open circles and triangles represent patients identified as outliers from severe or mild groups, respectively, as described in the text. The solid lines, dotted lines, and dashed lines are the estimated population regression lines for the young mild, older mild, and severe groups, respectively.

TABLE 1. CHARACTERISTICS OF STUDY GROUPS*

	Severe, 8–25 yr of Age (n = 256)	Younger Mild, 15–28 yr of Age (n = 304)	Older Mild, ≥ 29 yr of Age (n = 242)
Characteristics at time of enrollment			
Male, %	49.2	51.6	61.2 ^{§§}
Age at enrollment	16.1 ± 4.2	20.9 ± 4.0 [‡]	37.9 ± 5.3 ^{‡§}
Body mass index z-score	−1.28 ± 1.20	−0.11 ± 0.79 [‡]	−0.39 ± 1.08 ^{‡**}
FEV ₁ (% pred)	46.9 ± 17.3	91.7 ± 17.9 [‡]	50.4 ± 22.6 ^{‡††}
FVC (% pred)	62.8 ± 17.7	101.8 ± 15.3 [‡]	71.0 ± 21.6 ^{‡§}
FEF _{25–75} (% pred)	26.4 ± 20.0	75.7 ± 32.0 [‡]	25.7 ± 23.0 [‡]
FEV ₁ /FVC ratio	0.65 ± 0.12	0.78 ± 0.09 [‡]	0.58 ± 0.13 ^{‡§}
Number of PFTs per patient in previous 5 yr (range)	31.6 ± 23.9 (3–191)	18.3 ± 10.4 (2–72) [‡]	17.7 ± 12.5 (1–61) [‡]
Regressions of FEV ₁ (% pred) versus age [†]			
FEV ₁ slope (%pred/yr)	−3.6 ± 2.2	−1.1 ± 1.8 [‡]	−1.6 ± 1.1 ^{‡**}
FEV ₁ (% pred) at 15 yr of age	52.2 ± 14.9	98.3 ± 14.2 [‡]	86.7 ± 12.6 ^{‡§}
FEV ₁ (% pred) at 20 yr of age	34.2 ± 17.1	92.8 ± 14.7 [‡]	78.7 ± 12.1 ^{‡§}

Definition of abbreviation: PFT = pulmonary function test.

* Unless otherwise noted, table entries are mean ± SD. Between-group comparisons made using χ^2 tests and pair-wise comparisons in a one-way analysis of variance.

[†] Mixed model estimates obtained from mixed linear model (see Appendix 1).

[‡] Differs p < 0.0001 from severe group.

[§] Differs p < 0.01 from severe group.

^{§§} Differs p < 0.05 from younger mild group.

[‡] Differs p < 0.0001 from younger mild group.

^{**} Differs p < 0.01 from younger mild group.

^{††} Differs p < 0.05 from severe group.

from 0.5 for an index with no discriminating ability to 1.0 for an index able to discriminate perfectly (i.e., where there is no overlap in distribution between mild and severe groups).

One possible index of phenotype is the predicted probability of being severe from the logistic regression of disease severity on age and estimated FEV₁ (% pred) at enrollment (i.e., the method used to identify outliers in this study). However, because the severity groups are defined by this method, it achieves perfect separation by definition, and thus its performance cannot be directly assessed in this study. We focus on other indices, such as level of FEV₁ at a fixed age, patient-specific slope of FEV₁, and estimated survival percentile. Results are presented based on analysis of all 802 patients and for the subset of 553 patients who had 5 yr of pulmonary function measurements with at least one measurement in each year, to reduce the confounding influence of variability in length of follow-up (Table 2). Histograms of selected indices by severity group are displayed in Figures 3A–3D.

FEV₁ (% pred) at a fixed age. We examined the discriminating ability of the patient's estimated level of FEV₁ (% pred) at 5, 10, 15, 20, and 25 yr of age, obtained using mixed model and LS estimates of patient-specific slopes and FEV₁ levels (Table 2). Using mixed model estimates, optimal discrimination is achieved using estimated FEV₁ at around 20 yr of age (ROC area = 0.995). Estimated FEV₁ at the younger ages of 5 and 10 yr, which represent extrapolations back to young ages for many patients, were not as accurate in distinguishing between mild and severe groups. A histogram of the estimated FEV₁ at 20 yr of age (Figure 3A) indicates that there is little overlap between severe and mild groups with respect to this measure. As indicated by the ROC areas in Table 2, LS estimates of slopes or levels of FEV₁ (% pred) do not discriminate between groups as well as the mixed model estimates do. It is well known that LS slopes can be highly variable and imprecise for patients with few data points over a short time period (17–20). When restricted to subjects with 5 yr of data, the discriminating ability of LS estimates improved, but mixed model estimates still outperformed the LS estimates (Table 2).

Slope of FEV₁ (% pred). Areas under the ROC curve were 0.821 and 0.703 for the mixed model and LS estimates of FEV₁ slopes, respectively. Neither the mixed model (Figure 3B) nor LS estimates of FEV₁ slope (Figure 3C) were as good at discriminating between mild and severe patients as was the estimated level of FEV₁ at 15–25 yr of age.

Logistic regression on FEV₁ slope and intercept. Because each patient's regression line is fully characterized by its intercept and slope, another approach to developing a classification of mild versus severe patients is to fit a logistic regression of phenotype (severe vs. mild) using estimated FEV₁ intercept at age 0 and

TABLE 2. DISCRIMINATION BETWEEN SEVERITY GROUPS USING RECEIVER OPERATING CHARACTERISTIC CURVE AREAS

Index	Area under ROC Curve	
	All Patients (n = 802)	Patients with 5-yr Follow-up (n = 553)
Mixed model estimates		
FEV ₁ , 5 yr of age	0.722	0.689
FEV ₁ , 10 yr of age	0.880	0.860
FEV ₁ , 15 yr of age	0.980	0.975
FEV ₁ , 20 yr of age	0.995	0.994
FEV ₁ , 25 yr of age	0.987	0.985
FEV ₁ slope	0.821	0.822
FEV ₁ slope and intercept*	0.995	0.994
Least-squares estimates		
FEV ₁ , 5 yr of age	0.594	0.597
FEV ₁ , 10 yr of age	0.722	0.741
FEV ₁ , 15 yr of age	0.835	0.869
FEV ₁ , 20 yr of age	0.888	0.929
FEV ₁ , 25 yr of age	0.891	0.935
FEV ₁ slope	0.703	0.744
FEV ₁ slope and intercept*	0.894	0.937
Predicted survival percentile	0.995	0.991

Definition of abbreviation: ROC = receiver operating characteristic.

* Logistic regression of severity group on patient-specific FEV₁ slope and intercept.

Figure 3. Histograms of disease severity indices in severe and mild groups. Histograms are shown for (A) FEV₁ (% pred) at 20 yr of age, (B) slope of FEV₁ (% pred), (C) least-squares (LS) slope of FEV₁ (% pred), and (D) survival percentile from informative censoring model. Estimates in A and B are mixed model (empirical Bayes) estimates, using all available FEV₁ measurements for each subject. The x axis scales differ between B and C, reflecting the increased variability that is present in the LS slopes. Three outliers (two with slopes < 90%pred/yr and one with slope > 24%pred/yr) are excluded from C.

FEV₁ slope as predictors (Appendix 2). When this approach was used, the ROC area was 0.995 (Table 2). It can be shown (Appendix 2) that the ROC area of this approach using the intercept and slope is greater than or equal to the ROC area based on using estimated FEV₁ at any single fixed age. Also, provided the logistic regression coefficient of the intercept is nonzero (indicating that the FEV₁ intercept contributes to prediction in addition to the FEV₁ slope), using the intercept and slope to predict severity is equivalent to using the estimated level of FEV₁ at a single “optimal” age, estimated to be 19.6 yr in the GMS study (Appendix 2). The optimal age of 19.6 yr closely matches the empirically determined optimal age of 20 yr in Table 2, and the ROC area based on the FEV₁ at age 19.6 (0.995) is identical to that obtained using the intercept and slope.

Predicted age at death obtained from a model relating survival time to FEV₁. A model relating survival to FEV₁ (% pred), using estimates of parameters obtained by fitting it to an external dataset of 188 Δ F508 homozygous patients with CF (2), was applied to the GMS patient data to obtain a prediction of age at death. The prediction for each individual is expressed as a population percentile ranging from 0 to 1; thus, a predicted age at death equal to the median for the Δ F508 CF population would result in the percentile equaling 0.50, and lower (higher) values of the percentile represent worse (better) survival. Because the original parameter estimates from this model were obtained by fitting it to best yearly measures of FEV₁, we also used best yearly measures from the GMS data. The means and SDs of the estimated percentiles of predicted age at death were 0.37 ± 0.14 , 0.81 ± 0.09 , and 0.78 ± 0.06 for the severe, young mild, and older mild groups, respectively. Figure 3D shows distributions of the estimated survival percentiles. This approach provided an ROC curve area of 0.994, providing essentially the same discrimination between groups as the FEV₁ (% pred) at the optimal age of 19.6 yr.

Comparison of Estimates Obtained Using All versus Best Yearly FEV₁ Values

When analyzing longitudinal FEV₁ data in patients with CF, some authors (2) have analyzed the best (maximal) yearly measurements of FEV₁ on the grounds that these best yearly values are more representative of the patient’s usual condition, whereas the use of all available measurements may add variability and bias due to the inclusion of subnormal values obtained during acute illness episodes. On the other hand, subnormal FEV₁ values obtained during acute illness may relate to severity status, so it is not clear which approach is better. We therefore examined properties of estimates of FEV₁ slope and level at 20 yr of age, derived using the best (maximal) yearly FEV₁ per patient, and compared findings with results based on using all available FEV₁ values. Results are summarized in Table 3, where the comparison is restricted to patients who had measurements of FEV₁ in 5 consecutive years of data. For simplicity, we present comparisons for LS estimates only.

The average of the residual mean square error (Table 3), which is a measure of variability about each patient’s line, was smaller when the best yearly FEV₁ values were used. However, despite the fact that within-patient variability was reduced when using the best yearly FEV₁ values, the magnitudes of the SDs of LS slopes and estimates of FEV₁ at 20 yr of age were comparable for estimates based on all versus best yearly FEV₁, and in all cases the SDs were slightly larger when using the best yearly FEV₁. When examining the ability of LS estimates (intercept and slope) to discriminate between severe and mild groups, the ROC curve areas were 0.937 using all FEV₁ values and 0.894 using the best yearly FEV₁. Thus, in these data, analysis based on all FEV₁ values seemed to be preferable to analysis using the best yearly FEV₁.

When using the best yearly FEV₁ as compared with all FEV₁ values, the estimates of mean FEV₁ at a fixed age are higher.

TABLE 3. FEV₁ PARAMETERS ESTIMATED USING ALL DATA VERSUS BEST YEARLY (MAXIMAL) FEV₁ VALUES, RESTRICTED TO THE SUBSET OF PATIENTS HAVING 5 CONSECUTIVE YEARS OF DATA

	FEV ₁ Data Used	Severe (n = 182)	Young Mild (n = 229)	Older Mild (n = 142)
FEV ₁ (% pred) at 20 yr of age (least squares)	All	35.2 ± 22.9*	93.4 ± 19.3	77.5 ± 36.6
	Best yearly	44.4 ± 29.1	101.6 ± 20.0	80.0 ± 39.0
FEV ₁ slope (%pred/yr) (least squares)	All	-3.70 ± 3.02	-0.86 ± 2.87	-1.47 ± 2.08
	Best yearly	-3.73 ± 3.58	-0.41 ± 3.00	-1.42 ± 2.26
Mean residual mean square error about regression*	All	8.03	7.76	5.10
	Best yearly	5.90	5.49	3.85

* Table entries are the averages of the residual mean squares of the patient-specific least squares regressions of FEV₁ (% pred) on age in years.

This was more evident in the severe group than in the two mild groups. Similar results were seen when examining mixed model estimates (not shown). The group average slopes were generally comparable whether best yearly or all FEV₁ values were used. An exception was that the average LS slope for the young mild group based on best yearly FEV₁ values was approximately 50% less negative (less steep) as compared with the LS estimate using all FEV₁ values.

DISCUSSION

We demonstrate and empirically compare several approaches used to phenotype patients with CF according to severity of disease when longitudinal measures of lung function are available. The phenotypic summary measures derived from longitudinal data examined in this article have the following advantages over the use of a single FEV₁ measurement: (1) they automatically factor in the age of the patient, (2) summary measures typically are more precise than single measurements, and (3) they factor in the patient's current level and his/her previous rate of decline in FEV₁. Possible uses of these phenotypic disease severity indices in the context of gene modifier studies are (1) to provide a rigorous method to quantify or estimate pulmonary disease phenotype; (2) to define more extreme groups for comparison, such as the mildest versus the most severe of the patients; or (3) to use the index as a continuous measure of phenotype rather than dichotomizing patients into mild and severe groups. Along these lines, in a replication study performed as part of the GMS study (12), the estimated FEV₁ (% pred) at age 20 was used as a continuous phenotype and also dichotomized as < 68% versus ≥ 68% to examine the association of severity of pulmonary disease with the *TGFB1* codon 10 genotype.

The two approaches that best discriminated between severe and mild groups as defined in the GMS patients were (1) logistic regression using intercept and slope of FEV₁ (% pred; mixed model estimates) as predictors and (2) the predicted survival age percentile. ROC curve areas for these two approaches were essentially the same (0.995 and 0.994, respectively). Although we used intercept at birth and slope of FEV₁ in the model, the use of estimated FEV₁ at any other age along with slope provides an equivalent predictive model. We show that prediction of severity group using the logistic regression on slope and intercept is equivalent to using estimated FEV₁ at 19.6 yr of age as a single summary index and that the estimated level of FEV₁ at this "optimal" age outperforms FEV₁ estimated at other ages. Thus, the first approach can be simplified by using estimated FEV₁ at 20 yr of age as a single summary index. An advantage of this approach is its simplicity. A potential drawback is that the mixed model estimates of FEV₁ at a fixed age and slope are not independent of the group to which the subject was initially assigned

because the mixed model estimates are "shrunk" toward the mean of the group to which the patient was initially assigned. Although the survival percentile had the disadvantage of being more complex to calculate, it has several advantages: (1) it is estimated using only the subject's actual FEV₁ data and attained age and is thus independent of the initial group assignment and (2) it is calculated using parameters estimated externally to the sample of 802 subjects. The congruence of two very different methods (logistic regression based on slope and intercept, and survival percentile) in predicting severity was encouraging in the GMS and supports the validity of using either measure as a continuous severity measure.

Other summary indices, such as estimated level of FEV₁ (% pred) at 5, 10, 15, and 25 yr of age and FEV₁ slope, did not discriminate between severity groups as well as estimated FEV₁ at 20 yr of age or the estimated survival age percentile. The results indicate that the estimated slope of FEV₁ (% pred), whether estimated using the mixed model or LS, does not discriminate between mild and severe groups nearly as well as the patient's level of FEV₁ (% pred) at 20 yr of age, which is a function of the patient's estimated level of FEV₁ (intercept) and the slope. In an analysis using mixed models to compare intercepts and rates of decline in FEV₁ (% pred) of patients with CF, grouped by age at death, Corey and colleagues (3) similarly showed that patients who died earlier had lower intercepts and more negative slopes of FEV₁ (% pred), supporting the notion that severity is manifested through the initial level and the slope of FEV₁ (% pred).

The GMS sought to avoid the limitations of phenotyping on the basis of a single FEV₁ measurement; thus, longitudinal spirometry data were collected for the previous 5 yr on each subject, with the goal of better characterizing each patient's status using all the data. Study design and enrollment criteria were based on a number of clinical and practical considerations related to this approach, some of which are reinforced by these results. Only patients who had survived to the age of 8 yr were enrolled because of the need to obtain multiple measures of spirometry in previous years, with the expectation that many 8-yr-old patients would have 3 yr of data (i.e., at age 6–8 yr). Patients could be classified as severe at ages as young as 8 yr because having a low FEV₁ at a very young age is predictive of poor outcome (2, 3). The severe group extended only up to 25 yr of age because patients older than 25 yr would not be in the "worst" quartile of their birth cohort by virtue of surviving beyond 25 yr of age. For analogous reasons, patients who reached 34 yr of age were considered to have mild disease (i.e., were in the "best" quartile of their birth cohort) regardless of their level of pulmonary function because of their survival status. Finally, results presented here show that an individual patient's level of

FEV₁ at younger than 10 yr is not as reliable a predictor of severity as is the level at approximately 15 to 20 yr of age. The FEV₁ level at 20 yr of age reflects the patient's initial level and their rate of decline and can be reliably estimated only by following the patient to that age or by obtaining several years of data on younger patients. Similarly, Schluchter and colleagues (2) reported that, based on a joint model relating survival to FEV₁ longitudinal data in patients with CF, the highest correlation between age at death and the patient's level of FEV₁ (% pred) occurs in the range of 15 to 20 yr of age. For these reasons, the GMS did not attempt to classify a patient as "mild" before the age of 15 yr.

The mixed model estimates we evaluated were obtained from a model with severity group included in the model; thus, the intercept and slope estimates for each patient are shrunk toward the group mean to which the patient is initially assigned rather than toward the overall mean. An implication is that the estimate for an individual patient is influenced to some extent by the group to which the patient was initially assigned, particularly for patients with few points over a short age range. Thus, the mixed model estimates may seem to be somewhat better discriminators between mild and severe patients than they actually are, particularly when patients with short follow-up times are included. For example, among patients with FEV₁ measurements in five consecutive years of follow-up, the area under the ROC curve for a logistic regression using the individual estimates of slope and intercept obtained under a mixed model not including group as a fixed effect was 0.969, compared with an area of 0.994 for estimates obtained from a model with group included.

Mekus and colleagues (21) describe an alternative method for classifying severity of pulmonary disease in patients with CF by calculating age-specific percentiles of FEV₁ (% pred) in patients with CF, as determined from analysis of data in the European CF Registry. Using equations provided to us (F. Mekus-Stanke, personal communication), we calculated the percentile of FEV₁ (% pred) for GMS patients 30 yr of age or younger, the age range for which the calculations apply. Because this approach was originally developed using only a single FEV₁, we adapted it to our longitudinal data by using the mixed model estimate of level of FEV₁ (% pred) at the patient's final age and excluded patients older than 30 yr. The estimated ROC area for this index was 1.00, indicating complete separation between the groups. This is not unexpected because the mild and severe groups in the GMS study, by definition, show complete separation based on FEV₁ (% pred) at any given age, which is what the Mekus equation is based on. A possible drawback of the Mekus approach is that it does not consider survival. For example, two patients, 8 and 30 yr of age, who each have FEV₁ at the 50th percentile of their surviving cohorts, would be considered equivalent in severity, although the 30-yr-old patient has milder disease, when factoring in survival. Recently, Kulich and colleagues (22) published age- and gender-specific percentiles of FEV₁ (% pred) for patients with CF using data from the U.S. CF Foundation Patient Registry (1994–2001) for ages 6–40 yr in graphical form, which would allow similar percentiles to be calculated using data from patients with CF in the United States.

A limitation of the approach we have taken is that there is some circularity involved in defining severity and then comparing candidate severity measures against the groups so defined. The way severe and mild groups were initially defined precluded a "fair" comparison of the discriminating ability of the logistic regression using last predicted FEV₁ predicted and age as predictors, or of the Mekus approach, both of which result in 100% discrimination between groups in this dataset. However, the survival prediction, which is independent of the initial group assignment and estimated based on externally estimated param-

eters, yielded similar discrimination to the optimal approach using the patient's slope and intercept in a logistic regression model.

In conclusion, we have described an approach for enrolling and phenotyping patients with CF with severe and mild disease. Patients were enrolled initially based on age-specific cutoffs for their current FEV₁, and this classification of disease severity was confirmed after analysis of their longitudinal data in the 5 yr preceding enrollment. Only 18 of 820 (2%) of patients were deemed to be misclassified according to extremes of phenotype (severe vs. mild) based on the single initial FEV₁, when all longitudinal data were examined. Subsequent analyses suggest that the estimated FEV₁ (% pred) at 20 yr of age, calculated from longitudinal data, is an accurate discriminator between patients with severe and mild disease and can be linked to long-term outcome (i.e., duration of survival), which is the most important phenotypic outcome measure for testing the association of gene modifiers. Estimates of FEV₁ (% pred) at 20 yr of age from a longitudinal study may also be useful as a continuous index of disease severity, as evidenced by the approach recently used in the GMS (12).

Conflict of Interest Statement: None of the authors has a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

Acknowledgment: The authors thank Dr. Pamela Davis for support and for helpful input and feedback regarding the design of this study; Allison Handler, R.N., Colette Bucur, R.N., and Rhonda Pace, R.N., for assistance in coordination of the study and for assistance in obtaining and entering spirometric values into the database; Airong Xu, M.D., for establishing the Oracle-based relational database; and the Gene Modifier Study investigators who enrolled subjects and provided spirometric data (see listing in Appendix of Reference 12). This study was also supported by NIH grant RR00046 (to Principal Investigator Eugene P. Orringer) and NIH grant DK27651 (to Principal Investigator Pamela B. Davis).

References

1. Kerem E, Reisman J, Corey M, Canny GJ, Levison H. Prediction of mortality in patients with cystic fibrosis. *N Engl J Med* 1992;327:1244–1255.
2. Schluchter MD, Konstan MW, Davis PB. Jointly modelling the relationship between survival and pulmonary function in cystic fibrosis patients. *Stat Med* 2002;21:1271–1287.
3. Corey M, Edwards L, Levison H, Knowles M. Longitudinal analysis of pulmonary function decline in patients with cystic fibrosis. *J Pediatr* 1997;131:809–814.
4. Kerem E, Corey M, Kerem BS, Rommens J, Markiewicz D, Levison H, Tsui LC, Durie P. The relation between genotype and phenotype in cystic fibrosis: analysis of the most common mutation (delta F508). *N Engl J Med* 1990;323:1517–1522.
5. Newton-Chen C, Hirschhorn JN. Genetic association studies of complex traits: design and analysis issues. *Mutat Res* 2005;573:54–69.
6. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 2005;6:109–118.
7. Cutting GR. Modifier genetics: cystic fibrosis. *Annu Rev Genomics Hum Genet* 2005;6:237–260.
8. Sontag MK, Accurso FJ. Gene modifiers in pediatrics: application to cystic fibrosis. *Adv Pediatr* 2004;51:5–36.
9. Corey ML. Longitudinal studies in cystic fibrosis. In: Sturgess JM, editor. Perspectives in cystic fibrosis: proceedings of the 9th international congress on cystic fibrosis. Toronto, Ontario: Canadian Cystic Fibrosis Foundation; 1980. pp. 246–255.
10. Davis PB. Clinical pathophysiology and manifestations of lung disease. In: Yankaskas JR, Knowles MR, editors. Cystic fibrosis in adults. Philadelphia: Lippincott-Raven; 1999. pp. 45–67.
11. Gordon D, Finch SJ. Factors affecting statistical power in the detection of genetic association. *J Clin Invest* 2005;115:1408–1418.
12. Drumm ML, Konstan MW, Schluchter MD, Handler A, Pace R, Zou F, Zariwala M, Fargo D, Xu A, Dunn JM, *et al.* Genetic modifiers of lung disease in cystic fibrosis. *N Engl J Med* 2005;353:1443–1453.
13. U.S. Cystic Fibrosis Foundation. Patient registry 1999 annual data report. Bethesda, MD: Cystic Fibrosis Foundation; 2000.
14. Knudson RJ, Lebowitz MD, Holberg CJ, Burrows B. Changes in the normal maximal expiratory flow-volume curve with growth and aging. *Am Rev Respir Dis* 1983;127:725–734.

15. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36.
16. SAS Institute, Inc. SAS/STAT 9.1 User's guide. Cary, NC: SAS Institute, Inc.; 2004.
17. Burrows B, Lebowitz MD, Camilli AE, Knudson RJ. Longitudinal changes in FEV₁ in adults: methodologic considerations and findings in healthy nonsmokers. *Am Rev Respir Dis* 1986;133:974–980.
18. Schlesselman JJ. Planning a longitudinal study: II. Frequency of measurement and study duration. *J Chronic Dis* 1973;26:561–570.
19. Wu MC. Sample size for comparison of changes in the presence of right censoring caused by death, withdrawal, and staggered entry. *Control Clin Trials* 1988;9:32–46.
20. Davis PB, Byard PJ, Konstan MW. Identifying treatments that halt progression of pulmonary disease in cystic fibrosis. *Pediatr Res* 1997;41:161–165.
21. Mekus F, Laabs U, Veeze H, Tummler B. Genes in the vicinity of CFTR modulate the cystic fibrosis phenotype in highly concordant or discordant F508del homozygous sib pairs. *Hum Genet* 2003;112:1–11.
22. Kulich M, Rosenfeld M, Campbell J, Kronmal R, Gibson RL, Goss CH, Ramsey B. Disease-specific reference equations for lung function in patients with cystic fibrosis. *Am J Respir Crit Care Med* 2005;172:885–891.
23. Littell R, Milliken G, Stroup W, Wolfinger R. SAS System for Mixed Models. Cary, NC: SAS Institute; Inc., 1996.

APPENDIX 1. MIXED EFFECTS MODEL

Let y_{ij} be the j th measurement of FEV₁ (% pred) obtained on the i th subject, for $i = 1, \dots, n_i$. Define the group indicator variables $G_{1i} = 1$ if subject i is in the older mild group and 0 otherwise, and $G_{2i} = 1$ if subject i is in the severe group and 0 otherwise. The mixed model can be written as:

$$y_{ij} = (\alpha_0 + \alpha_1 G_{1i} + \alpha_2 G_{2i} + b_{i0}) + (\beta_0 + \beta_1 G_{1i} + \beta_2 G_{2i} + b_{i1}) + e_{ij}$$

where the α 's and β 's, . . . are unknown fixed effects parameters of the intercepts and slopes, b_{i0} and b_{i1} are random effects for the intercept and slope, and e_{ij} are independent normal errors with mean 0 and variance σ^2 . In this model, the mean intercepts

of FEV₁ are α_0 , $\alpha_0 + \alpha_1$, and $\alpha_0 + \alpha_2$ for the young mild, older mild, and severe groups, respectively, and mean slopes are β_0 , $\beta_0 + \beta_1$, and $\beta_0 + \beta_2$ for the same three respective groups. Mixed model estimates of individual slopes and intercepts (also called empirical Bayes estimates and Best Linear Unbiased Estimates) are obtained using SAS Proc Mixed as described in Littell and colleagues (23). Estimated mean intercepts ($\pm$ SE) for the young mild, older mild, and severe groups were 114.8 ± 2.7 , 110.7 ± 4.1 , and 106.2 ± 2.7 , respectively, and estimated mean slopes were -1.10 ± 0.14 , -1.60 ± 0.16 , and -3.60 ± 0.15 , respectively. Estimated variance components were: Var(intercept) = 1,485.0, Var(slope) = 4.017, Cov(intercept, slope) = -70.2, and $\sigma^2 = 63.3$.

APPENDIX 2. LOGISTIC REGRESSION OF SEVERITY GROUP ON FEV₁ INTERCEPT AND SLOPE

A logistic regression of severity (severe vs. mild) on the patient-specific estimated FEV₁ intercept (at age 0 yr) and slope can be expressed as Logit (P[severe]) = $\theta_1 + \theta_2(\text{intercept}) + \theta_3(\text{slope})$. By rearranging the right-hand side of the equation, for given values of the parameters θ_1 , θ_2 , and θ_3 (provided $\theta_2 \neq 0$), it can be expressed equivalently as: Logit P(severe) = $\theta_1 + \theta_2[\text{intercept} + \theta_3/\theta_2 (\text{slope})] = \theta_1 + \theta_2(\text{estimated FEV}_{1\%} \text{ predicted at age } \theta_3/\theta_2)$. Thus, when this model holds and $\theta_2 \neq 0$, it implies that an equivalent prediction is obtained using the single predictor defined as the patient's estimated FEV_{1%} predicted at age θ_3/θ_2 . That is, the ratio of parameters θ_3/θ_2 estimates the optimal age for predicting severity based on the level of FEV_{1%} predicted at that age.

This model was fit to the GMS data, yielding estimates (SEs) of the parameters: $\hat{\theta}_1 = 16.966$ (2.093), $\hat{\theta}_2 = -0.288$ (0.033), and $\hat{\theta}_3 = -5.640$ (0.612). The area under the ROC curve was 0.995. The optimal age was $\hat{\theta}_3/\hat{\theta}_2 = 19.6$ y, indicating that an equivalent predictive model is obtained by using the predicted FEV_{1%} predicted at age 19.6 yr as a single predictor in a logistic model.