31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Free-Knot Spline Model for Analysis of Pulmonary Function

Yong W. Lee, Jongwon Lee, *Member, IEEE*, Wha J. Yoo, Hyeong S. Yoo, and Warren J. Warwick

***Abstract*—The pulmonary function test (PFT) is used to evaluate and monitor respiratory function. The PFT is critical for the care of patients having cystic fibrosis (CF) and adjusting their clinical treatments. We analyzed the percent predicted value of forced expiratory volume in one second (FEV_1%) from PFT of CF patients collected four times a year from 1966. Longitudinal FEV_1% for each patient was fitted with linear free-knot spline (FKS) model. We explained as the time when the PFT trend changes. We classified the patients' pulmonary function trend using eight groups of FEV_1% based on the angle of linear FKS model. The overlapped majority of knots in groups located at 1978, 1991, and 1993 for worsening and at 1983, 1988, and 2000 for not worsening.**

I. Introduction

Many patients with cystic fibrosis (CF) develop chronic obstructive respiratory diseases. These patients have their pulmonary function tested at each clinic visit. These longitudinal data are sometimes overlooked. Commonly the recent test results are used to diagnose changes of the patient's respiratory condition. There is no treatment to cure the CF gene. Pulmonary illness is the predominant illness and cause of death. Many therapy methods have been studied and suggested for lung therapy. Mist tent [1-4] and high frequency chest compression (HFCC) [5-11] have shown that the lung diseases can be modified. Three companies manufacture and sale different HFCC systems. Each company is convinced that their system is the best.

Since the pulmonary function test (PFT) is a patient effort dependent measurement, the variance of measurement is inevitable. The doctors order for the PFT is performed by pulmonary function tester who does the test according to prescribed technology. Reviewing historical PFT data needs new methods to improve the evaluation. Even though the drop of lung function is common as CF patients grow older [12], changes of pulmonary function trend are often observed (Fig. 1.). Multi linear approximation with free-knot spline (FKS) method is successfully fitted when the change of

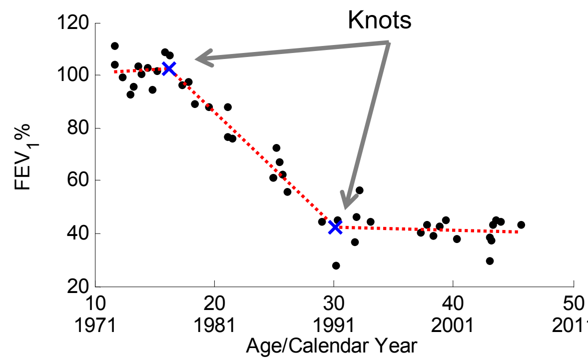


Fig. 1. The FEV_1% changes for CF patients during their lifetime. The trend of FEV_1% change can be approximated with couple of linear lines. Linear FKS model having two knots well matched to the longitudinal data. The location of knots may be explained with their age or calendar year which related to the treatments they took.

pulmonary function should be analyzed besides the variance of measurement.

Many doctors consider the forced expiratory volume in on second (FEV_1) is the best measurement to evaluate the lung function for all obstructive respiratory diseases. The patients FEV_1 is compared with normal population reference values; the ratio is generally used to generate the percent predicted values. Many doctors have suggested that their prediction model extracted from their populations at specific time and location can be used by others. The European Respiratory Society, the American Thoracic Society, and the Centers for Disease Control and Prevention have suggested using the largest studies to select the reference values for all pulmonary laboratories [13-16]. Still there is no standard reference that all use.

In this study, we propose a method to analyze the longitudinal change of percent predicted FEV_1 (FEV_1%) with their age and calendar year. We assume that the calendar year is appropriate parameter related to the therapies and clinical treatments for each CF patient.

II. Methods

A. Data Collection

We used PFT data from research database system of Minnesota Cystic Fibrosis Center, which is the oldest CF database system in US. The data from 1966 to 2007 containing 951 CF patients were analyzed. Our 951 patients are greater than the minimum required sample of 853 when $\alpha = 0.001$, number of predictors = 1, effect size = 0.02, and statistical power = 0.8. Every PFT result is affected by

Manuscript received April 7, 2009. This work was supported in part by the Defense of the Lung Project in U.S. Minnesota Cystic Fibrosis Center.
Y. W. Lee, and J. Lee are with the Department of Pediatrics, University of Minnesota, Minneapolis, MN 55455 USA (corresponding author to provide phone: 612-625-2672; e-mail: YWL: leeyongwan@gmail.com, JL: jongwona@gmail.com).
W. J. Yoo is with Physical Medicine and Rehabilitation Science Department, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: whayoo@gmail.com)
H. S. Yoo is with School of Computer Science and Engineering, Inha University, Incheon, 402-751 Korea
W. J. Warwick MD is with the Department of Pediatrics, University of Minnesota, Minneapolis, MN 55455 USA (phone: 612-624-7175; fax: 612-624-0413; e-mail: warwi001@umn.edu).

treatments each patient uses, thus, the regional difference in treatments will influence the sample distribution different from the population distribution.

PFT are performed more frequently while they are hospitalized. During the hospitalization, PFT data usually shows big improvement [12]. Because the hospitalization seldom affects the trend of pulmonary function, we excluded the hospitalization data in the analysis. New and young patients having less than 10 PFT measurements (227) were also excluded because the variance of PFT interferes the analysis for small data.

Our mixed reference model was used to calculate the percent predicted value of FEV_1. Our model provides the maximum third order polynomial equations having parameters: age, weight, and height.

B. Linear Free-Knot Spline Model

To generate multi linear model, the order of B-spline function was fixed to 2 that is a piecewise polynomial of degree 1. Maximum number of knots was limited to 2 which build 3 lines to fit the FEV_1%. We assume that the FEV_1% data, Y, can be modeled with a matrix form of B-spline function, $B(\kappa)$, with error, ε which has zero mean and bounded variance (1).

$$Y = B(\kappa)C(\kappa) + \varepsilon \quad (1)$$

From this equation (1), the coefficient vector, $C(\kappa)$, as the least square estimator is easily calculated by (2).

$$\hat{C}(\kappa) = \{B(\kappa)^{\mathrm{T}} B(\kappa)\}^{-1} B(\kappa)^{\mathrm{T}} Y \quad (2)$$

So the problem is to find the optimal knot vector which minimizing the error (3)

$$\hat{\kappa} = \underset{\kappa \in \Re^p}{\operatorname{argmin}} \left\| Y - B(\kappa)\hat{C}(\kappa) \right\|^2 \quad (3)$$

To solve this nonlinear minimization problem, stepwise knot-addition algorithm using Gauss-Newton algorithm was used [17]. The grid was assigned to the real data point distributed non-uniformly.

Since this method does not guarantee the global minimizer and the wide data length should be covered, we propose a procedure to determine the number of knots.

STEP1. Start with 2 knot model.

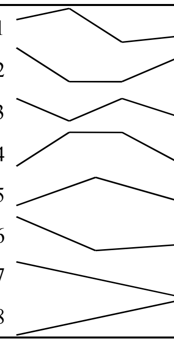
STEP2. If the estimated location of knot is next to the start or end data point or the angle of estimated lines is the outlier (4), from the distribution of all calculated angles, reduce the knot and check *STEP2* condition.

$$\dot{y} > Q1 + 2 \cdot IQR \, , \; \dot{y} < Q3 - 2 \cdot IQR \quad (4)$$

where $Q1$, $Q3$, and IQR are 25, 75 percentile and interquartile range respectively. Calculated range of outliers are [-22.4834, 21.5437] for 2 knot model, and [-27.0687, 29.8411] for one knot model.

The location of knots and angle of each line were calculated. The knot locations were interpreted with age and calendar year.

TABLE I
CLASSIFICATION OF PFT PATTERNS



Pattern	Angle Condition	Numbers	Age
1	$\dot{y}_1, \dot{y}_3 > \dot{y}_2$	195 (26.9%)	27.9
2	$\dot{y}_1 < \dot{y}_2 < \dot{y}_3$	18 (2.5%)	28.7
3	$\dot{y}_1, \dot{y}_3 < \dot{y}_2$	175 (24.2%)	27.4
4	$\dot{y}_1 > \dot{y}_2 > \dot{y}_3$	47 (6.5%)	28.8
5	$\dot{y}_1 > \dot{y}_2$	94 (13.0%)	21.5
6	$\dot{y}_1 < \dot{y}_2$	58 (8.0%)	23.8
7	$\dot{y}_1 < 0$	93 (12.8%)	25.9
8	$\dot{y}_1 \geq 0$	44 (6.1%)	17.1

Eight patterns were defined by angle condition and number of knots from linear FKS model. Minimum number of patients (18) was assigned to Pattern 2 which shows two consecutive improvement of pulmonary function. Number of knots is proportional to the average age of each pattern group. Decreasing patterns (Pattern 1, 3, 5, and 7) have higher number of patients than increasing patterns.

III. ANALYSIS/RESULTS

A. Classification

Eight groups were generated by the angle condition of each line and number of knots. Table I shows the pattern of each group and number of patients and their average age belongs to each group. Since the angle of line is compared relatively with the other angle, the pattern shapes on the Table I show average pattern. For example, Pattern 1 contains all positive angles which shows reverse pattern of Pattern 3. Greatest number of patients was assigned to Pattern 1 (26.9%), worsen and recovering shape. Second greatest one is Pattern 3 which shows reduced decreasing rate of pulmonary function in the middle of patients' records. Continuous improvement (Pattern 2) or drop (Pattern 4) had fewer patients. Both Pattern 5 and 6 have similar number of subjects and relatively young patients were assigned compared to 2 knot models (Pattern 1-4). Average age of Pattern 8 has minimum value 17.1.

TABLE II
MEAN AND STANDARD DEVIATION OF ANGLES

Pattern	$\dot{y}_1$		$\dot{y}_2$		$\dot{y}_3$	
	Mean	STD	Mean	STD	Mean	STD
1	3.95	6.27	-9.95	17.31	1.55	4.00
2	-7.50	6.37	-1.59	4.26	4.07	5.62
3	-6.63	5.98	7.32	8.66	-4.21	5.05
4	8.30	6.27	-0.21	3.72	-5.28	4.15
5	7.20	8.61	-5.97	6.16		
6	-9.03	8.19	3.57	6.14		
7	-3.10	3.72				
8	3.89	11.07				

The angle of each line has a normal distribution and their mean and standard deviation (STD) is shown in Table II.

B. FEV_1% by Age

Four major patterns, Pattern 1, 3, 5, and 6, having knots

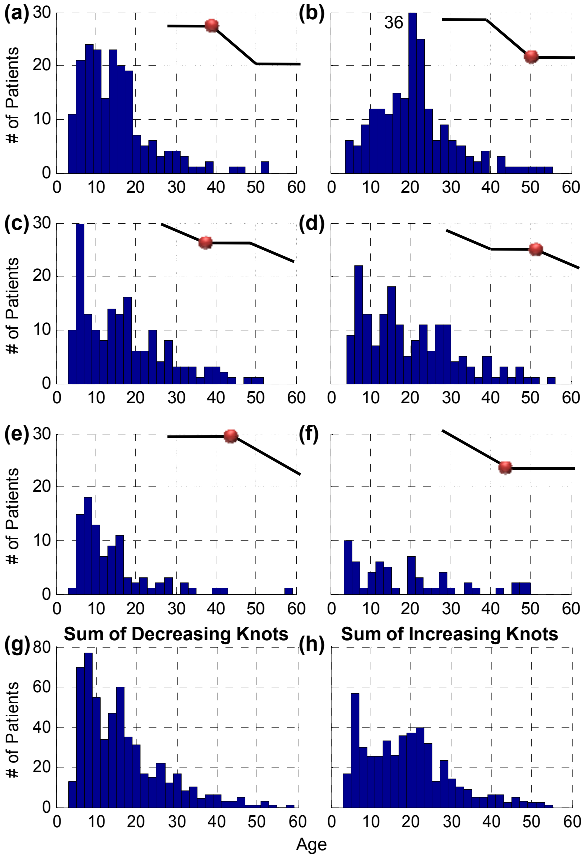


Fig. 2. Histogram (a) to (f) shows the age distribution of patients at a specific knot of PFT patterns. (g) and (h) are the sum of patients having decreasing and increasing knots respectively. Most of young patients below age 10 worsen their lung function. The decrease of PFT lessens or improves at early 20s.

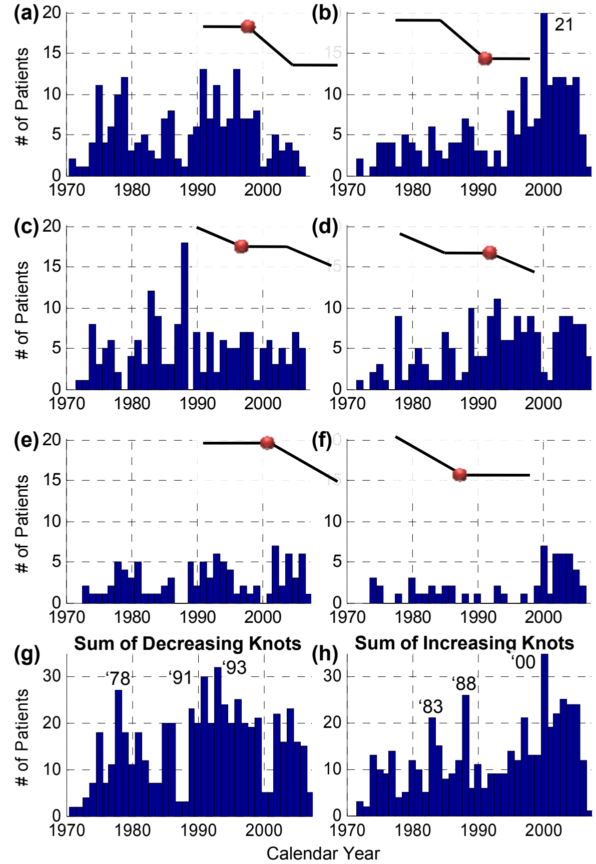


Fig. 3. Histogram (a) to (f) shows calendar year distribution of patients at a specific knot of PFT patterns. (g) and (h) are the sum of patients having decreasing and increasing knots respectively. Three maximum number of patients are at 1978, 1991, and 1993 for worsening case and 1983, 1988, and 2000 for reducing of PFT decrease or improving.

were investigated. Fig. 2. (a) and (b) show the distribution of patient's age for first and second knot of Pattern 1 respectively. Most of patients were around age 10 for first knot (a), age 20 for second knot (b). Pattern 3 is shown Fig. 2. (c) and (d) for each knot. These two knot locations have similar peaks below age 10. Patten 5 (e) has also peak below age 10; the patients' age distribution for Pattern 6 (f) is not clear because of the small number of samples.

We defined the decreasing and increasing knots for showing that the angle after knot decrease and increase from before respectively. So, Fig. 2. (a), (d), and (e) belong to decreasing knot and (b), (c), and (f) belong to increasing knot. The sum of patient number for decreasing and increasing knots is shown Fig. 2. (g) and (h). Median ages are 14.3 and 18.2 for (g) and (h) respectively.

C. *FEV_1% by Calendar Year*

Fig. 3. shows the histograms for patient distribution by tested calendar year. Only abscissa was converted to calendar year from Fig. 2.

It is not clear to see some distribution pattern when it is compared to age distribution. It is interesting that around early 1990s located the group of peaks of histogram (a) is the time that small number of patients was assigned. The opposite pattern, small for (a) and large for (b), also happens at 1988 and 2000. These patterns are same also for Fig. 3. (c) and (d) at early 1990s for decreasing knot and at 1988 and 2000 for increasing knot. It is not clear for (e) and (f) to see these patterns because of lack of patients. Fig. 3. (g) and (h) showing sum of decreasing and increasing knots clearly distinguish the pattern of the selected years.

IV. Discussion

We suggested a classification method and found the features of some classified groups. Since the limitation of the data, we only compared FEV_1% with age and measured calendar year. The linear FKS approximation successfully extracted features with the limited data. It seems that some environmental circumstances affected the decreasing knots in the selected years, 1978 and early 1990s, and the increasing knots 1983, 1988, and 2000. The PFT change by age can be explained with the life factor changes. Understanding PFT changes by age will require future studies of cultural items including parental care, financial and social independence, health insurance changes, and treatments (which chest therapy machine used, how used, how often, and how long). Year analysis may have influenced by the prevailed therapies in the year but it is not known which therapy is actually used most in each year. As a future work, the investigation of the therapy history will help to explain the pattern of years.

The mean life expectancy of CF patients has been increasing over 40 years from 4 to 37 today [18, 19]. That is the reason that most of age histograms have Rayleigh distribution. This will increase as life expectancy advances as new therapy methods are introduced, i.e. the center of age distribution will move to higher ages.

V. CONCLUSION

The longitudinal analysis of pulmonary function was done with linear FKS model plus the knot selection procedure. The median age of worsening lung function is 14.3 and of improving lung function is 18.2. The FEV_1% worsen in 1978, 1991, and 1993. The dropping speed of FEV_1% was reduced or pulmonary function was improved in 1983, 1988, and 2000.

REFERENCES

[1] L. W. Matthews, C. F. Doershuk and S. Spector, "Mist tent therapy of the obstructive pulmonary lesion of cystic fibrosis," *Pediatrics,* vol. 39, pp. 176-185, February 1, 1967. 1967

[2] M. A. Bureau, D. M. McDougall, P. H. Beaudry and M. M. Belmonte, "Late Effect of Nocturnal Mist Tent Therapy Related to the Severity of Airway Obstruction in Children With Cystic Fibrosis," vol. 61; Article, 1978.

[3] C. F. Doershuk, L. W. Matthews, C. T. Gillespie, M. D. Lough and S. Spector, "Evaluation of jet-type and ultrasonic nebulizers in mist tent therapy for cystic fibrosis," *Pediatrics,* vol. 41, pp. 723-732, April 1, 1968. 1968.

[4] B. M. Jakobsson, L. Hjelte and B. Nystr, "Low level of bacterial contamination of mist tents used in home treatment of cystic fibrosis patients," *Journal of Hospital Infection,* vol. 44, pp. 37-41, 2000.

[5] K. M. Chaisson, S. Walsh, Z. Simmons and R. L. Vender, "A clinical pilot study: High frequency chest wall oscillation airway clearance in patients with amyotrophic lateral sclerosis," vol. 7; Article, pp. 111, 2006.

[6] D. R. Hess, "The Evidence for Secretion Clearance Techniques," *Respir Care,* vol. 46, pp. 1276-1292, 2001.

[7] R. R. Kempainen, C. B. Williams, A. Hazelwood, B. K. Rubin and C. E. Milla, "Comparison of High-Frequency Chest Wall Oscillation With Differing Waveforms for Airway Clearance in Cystic Fibrosis," *Chest,* vol. 132, pp. 1227-1232, October 1, 2007. 2007.

[8] B. McCarren and J. A. Alison, "Physiological effects of vibration in subjects with cystic fibrosis," *Eur Respir J,* vol. 27, pp. 1204-1209, June 1, 2006. 2006.

[9] C. E. Milla, L. G. Hansen and W. J. Warwick, "Different Frequencies Should Be Prescribed for Different High Frequency Chest Compression Machines," *Biomedical Instrumentation & Technology,* vol. 40, pp. 319-324, 2006.

[10] W. J. Warwick and J. Braverman, "Thoracic oscillation, such as high frequency chest wall oscillators, provides oscillation directly to the chest wall." *Chronic Respiratory Disease,* vol. 4, pp. 51, February 1. 2007.

[11] T. Zucker, N. M. Skjodt and R. L. Jones, "Effects of High-Frequency Chest Wall Oscillation on Pleural Pressure and Oscillated Flow," *Biomedical Instrumentation & Technology,* vol. 42, pp. 485-491, 2008.

[12] Y. W. Lee, J. Lee, H. S. Yoo and W. J. Warwick, " Pulmonary function analysis of patients with cystic fibrosis," *Pediatric Pulmonology,* vol. 43, pp. 371, 2008.

[13] J. E. Cotes, D. J. Chinn, P. H. Quanjer, J. Roca and J. C. Yernault, "Standardization of the measurement of transfer factor (diffusing capacity). Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society." *Eur Respir J,* vol. 6, pp. 41-52, 1993.

[14] American Thoracic Society, "Lung function testing: selection of reference values and interpretative strategies." *Am Rev Respir Dis,* vol. 144, pp. 1202-1218, 1991.

[15] American Thoracic Society, "Single-breath carbon monoxide diffusing capacity (transfer factor). Recommendations for a standard technique--1995 update." *Am J Respir Crit Care Med,* vol. 263, pp. 2185-2198, 1995.

[16] J. L. Hankinson, J. R. Odencrantz and K. B. Fedan, "Spirometric reference values from a sample of the general U.S. population," *American Journal of Respiratory and Critical Care Medicine,* vol. 159, pp. 179-87, Jan. 1999.

[17] D. Gervini, "Free-knot spline smoothing for functional data," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 68, pp. 671-687, 2006.

[18] "Cystic fibrosis foundation 2007 annual report," Cystic Fibrosis Foundation, 2007.

[19] M. Corey and V. Farewell, "Determinants of Mortality from Cystic Fibrosis in Canada, 1970-1989," *Am. J. Epidemiol. ; Am. J. Epidemiol.,* vol. 143, pp. 1007-1017, May 15. 1996.