Eur Respir J 1999; 13: 552–559
Printed in UK – all rights reserved

European Respiratory Journal
ISSN 0903-1936

Construct and longitudinal validity of a modified Huang clinical scoring system in adult cystic fibrosis patients

E. Matouk, R.H. Ghezzo, J. Gruber, R. Hidvegi, K. Gray-Donald

Construct and longitudinal validity of a modified Huang clinical scoring system in adult cystic fibrosis patients. E. Matouk, R.H. Ghezzo, J. Gruber, R. Hidvegi, K. Gray-Donald. ©ERS Journals Ltd 1999.

ABSTRACT: This study reports on the evaluation of a modified Huang scoring system in adult cystic fibrosis patients for construct and longitudinal validity.

Two studies were performed. In the first study, the scoring system was applied to 59 adult cystic fibrosis patients prospectively followed at the Montreal Chest Institute. The total score and all the subscores distinguished between patients with the expected mild degree of disease severity seen in patients colonized with only *Staphylococcus aureus*, compared to the more advanced disease severity seen in patients colonized with *Pseudomonas aeruginosa* or multiple resistant pseudomonads. The relationship between disease severity assessed by forced expiratory volume in one second per cent predicted and the nonpulmonary function subscores was significant and linear (for the radiological subscore, r^2=0.694, p<0.0001) and curvilinear (for the clinical and complications subscores, r^2=0.622, p=0.0192 and r^2=0.508, p=0.0009 respectively).

In the second study, 20 patients retrospectively recorded were added to the prospective group. There was a good association between changes in nonpulmonary function subscores and changes in spirometry over a mean follow-up period of 779±204 days, at all levels of disease severity. The contribution of changes in clinical and complications subscores to the changes in total score became progressively more significant with more advanced disease severity.

In conclusion, significant evidence for the construct validity of the scoring system as a discriminative instrument and for the longitudinal validity as an evaluative instrument was demonstrated. It may prove of value in assessing outcome of therapeutic interventions in clinical trials in patients with cystic fibrosis.

Eur Respir J 1999; 13: 552–559.

Respiratory Division, Montreal Chest Institute and Research Centre, McGill University, Montreal, Québec, Canada.

Correspondence: E. Matouk
Montreal Chest Institute
3650 St Urbain Street
Montreal
Quebec
Canada, H2X 2P4
Fax: 514 8432083

Keywords: Cystic fibrosis
construct validity
longitudinal validity
questionnaires
severity of illness index

Received: January 15 1998
Accepted after revision October 6 1998

Cystic fibrosis (CF) is the most common serious autosomal recessive disease affecting the Caucasian population [1]. It is characterized by a suppurative, inflammatory, obstructive and progressive pulmonary disease [2]. It is heterogeneous in its clinical presentation at different ages, and in the severity of involvement and rate of progression in affected organs [3].

The improvement in survival, increase in the number of adult cystic fibrosis patients and the advances in treatment modalities, have created a need for the development and evaluation of a clinical scoring system for adult CF patients [4, 5].

A clinical scoring system for these patients has been modified from the one proposed by HUANG *et al.* [6]. The total score represents the sum of the clinical, radiological and pulmonary function subscores, minus the complications subscore. It ranges between 0 (worse health) and 100 points (best health). A prior study has demonstrated a high degree of internal consistency reliability and predictive validity of the computed total score [7].

The present paper describes two studies. Study 1 deals with the evaluation of the construct validity of the scoring system. It refers to whether the scoring system measures what it claims to measure, *i.e.* CF health status at a point in time [8, 9]. Study 2 evaluates the scoring system for changes over an extended period of time. This longitudinal study relates changes in the scores to changes in other accepted measures of disease severity such as forced expiratory volume in one second (FEV_1) per cent predicted over a period of approximately 2 yrs [8, 9].

Subjects and methods

Patient population

Data for the present study came from a database of 109 patients followed at the Montreal Chest Institute (MCI) between 1967 and 1994. Ninety-nine patients had been transferred from the Montreal Children's Hospital and 10 patients were diagnosed for the first time after the age of 18 yrs-old. All patients had a positive sweat chloride test.

The modified scoring system was first developed in 1991 and used prospectively (study 1) on 59 patients during a period of follow-up between December 1991 and July 1994 (31 months).

In the second study, 20 patients were added from the retrospective pool of 50 patients. These patients were

followed by the principal investigator between 1982 and 1991 and were of similar mean age and degree of disease severity compared to those most severely affected in the prospective group at the last recorded visit. They were added in order to increase the number of patients with advanced degree of disease severity in this study. Two patients who left the prospective group were excluded from the longitudinal analysis because the period of follow-up was too short.

Data collection

In Study 1, subjective data on cough, sputum, dyspnoea and appetite were obtained from each patient at each visit using a 1 page multi-item questionnaire completed by the patient. Objective data obtained at each visit included respiratory rate, pulse rate, temperature, height, weight and physical examination. Spirometry (forced expiratory volume in one second (FEV_1) % pred, forced vital capacity (FVC) % pred, FEV_1/FVC ratio, forced midexpiratory flow rate ($FEF_{25–75\%}$) % pred, forced midexpiratory flow ($FEF_{50\%}$) % pred), was performed according to standards established by the American Thoracic Society [10]. The best of two reproducible efforts was recorded. With the exception of FEV_1/FVC ratio, they were each expressed as a percentage of the predicted normal value for the patient's age, sex, and height [11]. Arterial oxygen saturation (S_{a,O_2}) was obtained at each visit by fingertip pulse oximetry (Nellcor N-100, Nellcor Puritan Bennett, Pleasanton, CA, USA). Arterial blood gases were obtained if S_{a,O_2} <90% and or when clinically indicated (particularly in patients with FEV_1 % pred <34%). Previous studies by WAGENER *et al.* [12] have shown that it was rare for a patient with a FEV_1 % pred ≥34% to have a high arterial carbon dioxide tension (P_{a,CO_2}) and 47.9% of the patients with CF with an FEV_1 % pred <34% were carbon dioxide retainers [12]. These data were used to compute the respiratory failure item score of the scoring system.

Data used in the first study came from the prospective group of patients at the last recorded visit while they were considered clinically stable, *i.e.* no pulmonary exacerbation recorded for at least 4 weeks prior to the reference visit. Pulmonary exacerbation was defined as a change in the patient status resulting in a change in treatment (*i.v.* antibiotics or *p.o.* antibiotics such as ciprofloxacin, trimethoprim-sulphamethoxazole and cephalexin or inhaled tobramycin + *p.o.* antibiotics) [7].

Chest radiographs were scored by a staff radiologist (R. Hidvegi) who had no knowledge of the other components of the score. The timing of the chest radiographs used in the scoring at the last recorded visit was either at the same visit or between 1 to 2 months from that visit while the patient was considered clinically stable.

In study 2, the retrospective data that were added to the prospective group were obtained from the 20 patients seen between 1982 and 1991, through review of chest radiographs and medical records. The same data were readily obtained from the outpatient visits and in some cases from hospitalizations. The subjective data (cough, sputum, dyspnoea and appetite) were part of the routine history documented in the medical records of the patients. Arterial blood gases, obtained during periods of clinical stability to assess oxygenation and alveolar ventilation, were used to compute the respiratory failure item score of the scoring system. Data from visits of patients in acute exacerbations were not included in the analyses. Data used in the second study came from this combined group of patients at the last recorded visit and at a visit recorded approximately 2 yrs earlier while the patients were considered clinically stable.

Patients who had undergone lung transplantation were included in the cross-sectional analysis. The date of scoring was at the last visit prior to the transplant, while the patient was considered clinically stable.

A relational database program was used for data entry and scoring.

Statistical analysis

Study 1: construct validity. The scoring system for construct validity as a discriminative instrument was evaluated first. There is no simple method for evaluating the construct validity of a scoring system. None of the construct validation procedures alone offers definitive proof of validity and all suffer from logical and practical limitations [13]. However, it may be assessed by examining the ability of the scoring system to distinguish between groups of patients with expected differences in disease severity [8, 9]. For example, patients colonized with *Staphylococcus aureus* alone were expected to have milder disease severity than those colonized with *Pseudomonas aeruginosa* [1, 14–16]; and those colonized with *Burkholderia cepacia* or multiple resistant pseudomonads (defined as the presence of pseudomonas organisms resistant to all antibiotics included in all the different classes on at least two samples of sputum close to the scoring visit) were expected to have worst disease severity compared to the other two groups [17]. The Mann–Whitney U-test was used to assess whether the total score and subscores differed significantly among these groups. Significance was accepted at $p \leq 0.05$.

Construct validity was also assessed by examining the correlation between the nonpulmonary function subscores and other clinical measures of disease severity such as FEV_1 % pred [8, 9]. Linear regression and second order polynomial regression plots were computed where indicated. Significance was accepted at $p \leq 0.05$.

Study 2: change over approximately 2 yrs. In this observational longitudinal study, the scoring system was evaluated for changes over an extended period of time, approximately 2 yrs. This was done overall and in three groups of patients divided according to the degree of disease severity defined by the total score at last recorded visit. The first group of patients (n=33) with total score <35 represented the most severely ill patients. It included patients, who according to current practice, would be on the transplantation list. The second group of patients (n=25) with total score ≥35 and <60 represented those with moderate disease. The third group of patients (n=19) with total score ≥60 represented the asymptomatic patients and those with mild disease.

The degree of change was evaluated by estimating the mean absolute change in total score, subscores and spirometry. This was carrried out by subtracting the corresponding value at last recorded visit from the corresponding value at a visit approximately 2 yrs earlier, while the patient was considered clinically stable.

Table 1. – Patient demographics and characteristics at last recorded visit (Study 1; n=59 prospectively followed up)

Sex male/female	31/28
Age at last visit yrs	28.3±6 (18–42)
FEV_1 % pred	49±28 (9–112)
FVC % pred	68±29 (11–112)
IBW % pred	91±16 (68–151)

Values are shown as mean±SD, with range in parentheses. FEV_1: forced expiratory volume in one second; FVC: forced vital capacity; IBW: ideal body weight.

Then, the contribution of changes in individual item scores and subscores to changes in total score was assessed. Pearson correlation coefficients (r) were used to assess the correlations. In view of the multiple variables and the multiple possible comparisons, the $p<0.05$ end point for significance was calculated using Bonferroni's adjustment for multiple comparisons. For the entire group of patients, n=77 and 250 correlations, statistical significance was accepted at $r>0.342$. For the group total score <35, with n=33 and 250 correlations, statistical significance was accepted at $r>0.487$. For the group total score ≥35 and <60, n=25 and 250 correlations, statistical significance was accepted at $r>0.54$. For the group total score ≥60, n=19 and 250 correlations, statistical significance was accepted at $r>0.58$. In view of the multiple comparisons, these correlations should be interpreted only as relative measures of an association between the item or subscore and total score.

Finally, the correlation was assessed between the mean absolute change in the nonpulmonary function scores and mean absolute change in FEV_1 % pred and FVC % pred, as well as the mean per cent change in $FEF_{25–75\%}$ % pred and $FEF_{50\%}$ % pred [18]. The coefficient of determination (r^2) expressed the proportion of the variance in nonpulmonary function scores (dependent variable) attributable to the variance in the other variables.

Results

Study 1

Patient population, visits and frequency distribution. Table 1 shows the characteristics of the 59 patients in the first study at the time of the last recorded visit. Of the 59 patients in the prospective group, the scoring was carried out at every visit during the follow-up period 1991–1994, an average of 19 visits per patient. Fifty-one patients were alive and still being followed. One is alive 4.5 yrs post-lung transplantation and two patients left the clinic and were alive at the last visit. Five patients died during the second study period, of whom one died 12 months post-transplantation. In the 20 patients of the retrospective group added to the second study, the scoring was carried out on an average of four visits per patient during the second study period. They all died during this period.

As shown in figure 1, the distribution of disease severity in the patient population in study 1, using the total score, was more homogenous when compared to the more skewed distribution using FEV_1 % pred. The total score histogram suggested a greater discriminatory power of this measuring instrument as it enabled the detection of intermediate health states [19]. A prior study has demonstrated that the discriminatory power of the total score was genuine in terms of its predictive validity [7].

Construct validity. Table 2 lists the patient characteristics and scores in the prospective group of patients according to the organism consistently recovered on sputum culture for the 6 months preceding the last recorded visit. As expected, patients colonized with *S. aureus* alone had a significantly higher mean total score compared to the patients colonized with *P. aeruginosa* (mucoid or nonmucoid). This was owing to significantly higher mean clinical, radiological, pulmonary function subscores and a lower mean complications subscore. Furthermore, patients colonized with multiple resistant pseumonads had a significantly lower mean total score compared to those colonized with *P. aeruginosa* (mucoid or nonmucoid). This was owing to a

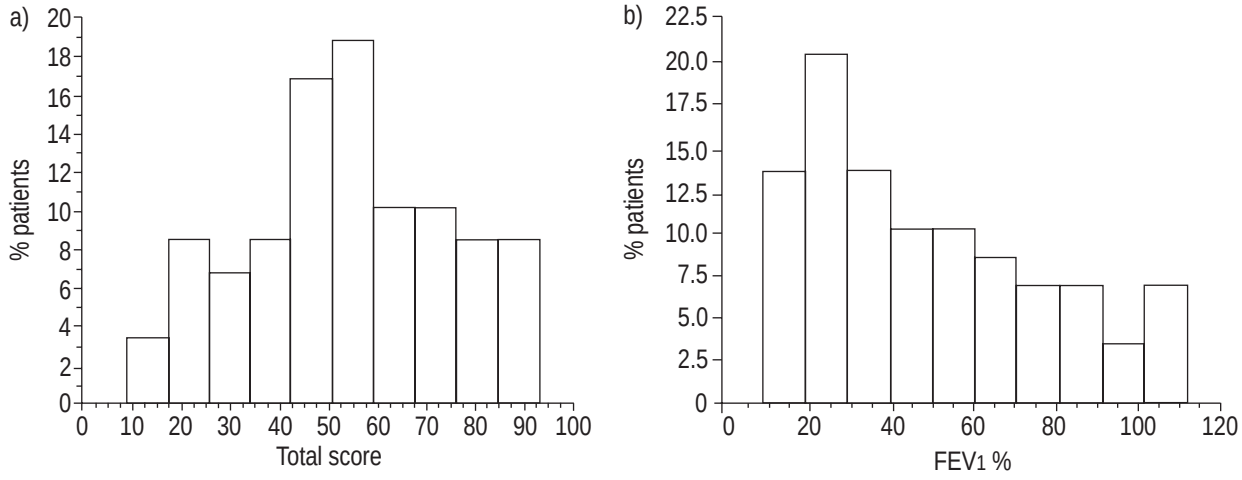

Fig. 1. – Frequency distribution of disease severity using a) total score and b) forced expiratory volume in one second (FEV_1) per cent predicted (study 1). The ordinate represents the percentage of patients with a given value.

Table 2. – Patient characteristics and scores according to the predominant organism on sputum culture

	Predominant organism				
	Staphylococcus aureus	*Pseudomonas aeruginosa* nonmucoid and mucoid	p-value*	Multi-resistant pseudomonas	p-value†
Patients n	11	43		5	
Sex male/female	8/3	20/23		3/2	
Age at last visit yrs	28.5±8 (18–42)	28.2±5.6 (18–39)		28.8±4.5 (21–32)	0.9328
Clinical subscore	41±4 (34–48)	31.5±6 (16–43)	<0.0001	21.8±8.2 (10–32)	0.0108‡
Radiological subscore	20±2 (14–23)	15±3 (11–23)	<0.0001	11.6±3 (8–16)	0.0321‡
PFT subscore	21±5.6 (8–25)	12.2±5 (5–25)	0.0003	8.8±4 (6–16)	0.0659
Complications subscore	2.3±2.8 (0–7)	8.7±5 (0–22)	<0.0001	14.4±8 (4–23)	0.1127
Total score	80±9.5 (59–93)	50±16 (12–91)	<0.0001	27.8±22 (9–60)	0.0259

Values are shown as means±SD, with range in parentheses. PFT: pulmonary function test. *: p-Value, Mann–Whitney U-test (*S. aureus versus P. aeruginosa*), all statistically significant; †: p-value, Mann–Whitney U-test (*P. aeruginosa versus* multi-resistant pseudomonas), ‡: statistically significant.

significantly lower mean clinical and radiological subscore and no significant differences in pulmonary function or complications subscores.

In this group of patients, the relationship was also examined between the nonpulmonary function subscores and disease severity assessed by FEV1 % pred. A highly significant linear correlation was seen between FEV1 % pred and the radiological subscore (r^2=0.694, p<0.0001). The relationship between FEV1 % pred and clinical as well as complications subscores was significant and curvilinear (r^2=0.622, p=0.0192 and r^2=0.508, p=0.0009 respectively). As shown in figure 2, the relationship between FEV1 % pred and the combined clinical/complications subscores was highly significant and curvilinear (r^2=0.66, p=0.0008).

All the curvilinear relationships were characterized by an initial plateau, followed by an accelerated decline. The threshold point at which the accelerated decline began corresponded to an FEV1 % pred of 60%.

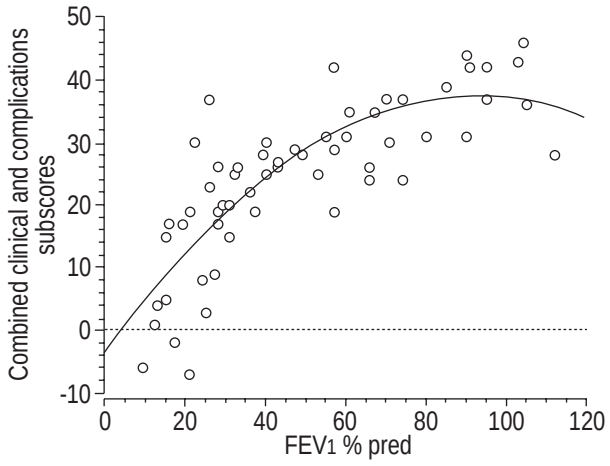

Fig. 2. – Correlation between combined clinical and complications subscores and forced expiratory volume in one second (FEV1) per cent predicted. p =0.0008, r^2 =0.660.

Study 2: Changes over approximately 2 yrs (longitudinal validity)

Table 3 shows the descriptive statistics of the three groups of patients, stratified according to disease severity defined by total score at last recorded visit. The characteristics of the 20 patients from the retrospective pool of 50 who were added in this study to the severe group of patients in the prospective cohort, were as follows: male (10 patients), female (10 patients), age at last visit (mean±SD: 26.6±6 yrs, range 18–41), FEV1 % pred (15±4%, range 8–25.5), FVC % pred (24±7%, range 10–42), and ideal body weight % pred (74±9%, range 59–98). Table 4 shows the degree of the changes in total score, subscores and spirometry overall and in the three individual groups of patients over a mean follow-up of 779±204 days. Table 5 shows the correlations of changes in individual item scores and subscores with changes in total score overall and at different levels of disease severity defined by total score at the last visit.

In the group total score <35 (n=33, duration of follow up (mean±SD) 787±253 days), the most predictive variables significantly contributing to the change in total score were complications subscore (mainly respiratory failure), clinical subscore (respiratory rate combined score, dyspnoea, weight, appetite and physical examination) and pulmonary function subscore (mainly FVC score).

In the group total score ≥35 and <60 (n=25, duration of follow up (mean±SD) 769 days ±144 days), the most predictive variables significantly contributing to the change in total score were clinical subscore (cough, respiratory rate combined score, dyspnoea and sputum scores), complications subscore, (pulmonary exacerbation and respiratory failure scores) and pulmonary function subscore (FVC, FEV1, FEF25–75% and FEF50% scores). In the group total score ≥60 (n=19, Duration of follow up (mean±SD) 780–187 days), the most predictive variables significantly contributing to the change in total score were clinical subscore (weight change score and a trend with culture and respiratory rate combined scores), pulmonary function subscore (FEF50% and a trend with FEF25–75%), radiological

Table 3. – Descriptive statistics of the three groups of patients defined by the level of total score (study 2)

Variable	Severe group Total score <35	Moderate group Total score ≥35 and <60	Mild group Total score ≥60
Patients n	33	25	19
Age at last visit yrs	27.2±5.7 (18–39)	28.5±5.7 (19–42)	26.9±6 (18–41)
Age at diagnosis months	33±49.7 (1–192)	45±96.7 (1–408)	66.7±96.7 (1–396)
FEV1 % pred	16.2±4.7 (8–27)	38.2±13.7 (19–74)	76.3±19.9 (47–112)
FVC % pred	28.9±8 (10–43)	59.7±17.8 (29–93)	92.7±19 (56–120)
FEV1/FVC Ratio	52.3±11 (36–83)	55.6±11.8 (40–82)	68.9±9.1 (50–86)
FEF25–75% % pred	8.2±2.8 (3–15)	15.3±9.5 (0.77–37)	51.5±30.8 (15–132)
FEF50% % pred	7.5±3.7 (2–15)	16.2±11.9 (0.86–47)	56.1±32.5 (24–146)
IBW % pred	79±13.7 (59–125)	83.6±8.9 (68–107)	98.8±17.1 (79–151)

Values are shown as means±SD, with range in parentheses. FEV1: forced expiratory volume in one second; FVC: forced vital capacity; FEF25–75%: forced midexpiratory flow rate; FEF50%: forced midexpiratory flow; IBW: ideal body weight.

subscore (air trapping, nodular cystic changes, and a trend with linear markings and general impression scores), and complications subscore (mainly pulmonary exacerbation score).

The variations between patients in the degree of changes in the measured variables were sufficient to allow testing for associations between changes in total score excluding pulmonary function subscore, and changes in spirometry over a mean follow-up period of 779±204 days (table 6). In each correlation, the residuals were normally distributed around a linear regression. When all the three groups of patients were included, the changes in total score excluding pulmonary function subscore significantly correlated with changes in FEV1 % pred (r^2=0.254, p<0.0001) and FVC % pred (r^2=0.437, p<0.0001) suggesting the longitudinal validity of these scores.

In patients with a total score <35, the changes in total score excluding pulmonary function subscore were significantly correlated with changes in FEV1 % pred (r^2=0.432, p<0.0001) and FVC % pred (r^2=0.386, p=0.0001). In patients with a total score ≥35 and <60, the changes in total score excluding pulmonary function subscore were significantly correlated with changes in FEV1 % (r^2=0.301, p=0.0045) and FVC % pred (r^2=0.434, p=0.0003). In patients with total score ≥60, the changes in total score excluding pulmonary function subscore were significantly correlated with changes in FVC % pred (r^2=0.227, p=0.039), FEF25–75% % pred (r^2=0.21, p=0.0482), FEF50% % pred (r^2=0.323, p=0.011) and a trend with changes in FEV1 % pred (r^2=0.203, p=0.053).

Discussion

A proposed scoring system was evaluated for construct validity as a discriminative instrument and for longitudinal validity as an evaluative instrument measuring changes over time at different levels of disease severity.

Table 4. – Degree of change in spirometry and scores at different levels of disease severity over an average follow-up period of 779±204 days

	Severe group Total score <35	Moderate group Total score ≥35 and <60	Mid group Total score ≥60	All groups
FEV1 L†	-0.328±0.24	-0.314±0.43	-0.159±0.374	-0.282±0.347
FEV1 % pred†	-9.3±7.8	-10±13.4	-3.9±12.2	-8.2±11.1
FVC L†	-0.707±0.5	-0.298±0.53	-0.054±0.51	-0.413±0.572
FVC % pred†	-17.5±12.2	-9.4±14.2	-2.7±15.7	-11.2±14.9
FEV1/FVC ratio†	1.8±9.9	-4.2±8.2	-3.3±6.3	-1.4±8.9
FEF25–75% % pred*	-0.26±0.31	-31.2±31.4	17.3±20.2	-7.8±12.2
FEF50% % pred*	-0.21±0.35	-33.3±31	-13.9±16.7	-7.2±11.5
Total score†	-22.45±11	-13.56±10.73	-8.42±7.96	-16.04±11.6
Clinical subscore†	-7.5±6	-5.2±5.1	-3.57±3.4	-5.82±5.4
Radiological subscore†	-1±1.6	-1.48±1.87	-0.68±2.7	-1.1±2
PFT subscore†	-2±2	-2.6±3.40	-1.47±2.1	-2.12±2.7
Complications†	-12±5	-3.72±5	-2.7±3.4	-6.99±6.5
Total score minus PFT subscore†	-20±10	-10.96±8.9	-6.9±6.9	-13.92±10.4

Values are shown as means±SD. FEV1: forced expiratory volume in one second; FVC: forced vital capacity; FEF25–75%: forced midexpiratory flow rate; FEF50%: forced midexpiratory flow; PFT: pulmonary function test; *: mean per cent change; †: mean absolute change.

Table 5. – Correlations of change (r-values) in individual item scores and subscores with change in total score over a mean follow-up period of 779±204 days

	Patient group (total score range)			
	<35	≥35–<60	≥60	All
Patients n	33	25	19	77
Average change in total score	-22.45	-13.56	-8.42	-16
Pearson correlation coefficients				
Weight	0.533*	0.491	-0.01	0.517*
Weight change	0.394	0.454	0.748*	0.363*
Dyspnoea	0.561*	0.637*	0.2	0.636*
Cough	0.237	0.682*	0.181	0.445*
Sputum	0.415	0.617*	0.18	0.4*
Physical examination	0.488*	0.404	0.019	0.423*
Respiratory rate combined score	0.651*	0.662*	0.449	0.451*
Culture	0.319	0.335	0.486	0.322
Appetite	0.532*	0.244	0.343	0.493*
General condition	0.441	0.518	0.322	0.503
Clinical subscore	0.751*	0.865*	0.767*	0.799*
Air trapping	0.195	-0.055	0.655*	0.18
Linear markings	0.128	0.061	0.52	0.106
Nodular cyst	0.177	0.143	0.649*	0.116
Parenchymal	0.116	0.254	0.214	0.263
General impression	0.343	0.189	0.489	0.303
Radiological subscore	0.332	0.173	0.576*	0.304
FVC	0.525*	0.696*	0.461	0.635*
FEV_1	0.374	0.632*	0.372	0.534*
FEV_1/FVC	0.153	0.266	-0.064	0.043
$FEF_{25–75\%}$	0.275	0.547*	0.522	0.286
$FEF_{50\%}$	0.283	0.543*	0.632*	0.327
Pulmonary function subscore	0.488*	0.671*	0.61*	0.524*
Pulmonary exacerabation	0.429	0.694*	0.734*	0.551*
Pneumothorax	0.099	-0.066	–	0.177
Haemoptysis	-0.01	0.08	0.024	0.069
Respiratory failure	0.623*	0.612*	–	0.643*
Cardiac enlargement	0.286	–	–	0.373*
Pulmonary surgery	0.177	–	–	0.171
Complications subscore	0.781*	0.783*	0.722*	0.824*

r: Pearson correlation coefficient. *: statistically significant using the Bonferroni adjustment for multiple comparisons. FVC: forced vital capacity; FEV_1: forced expiratory volume in one second; $FEF_{25–75\%}$: forced midexpiratory flow rate; $FEF_{50\%}$: forced midexpiratory flow.

The construct validity of a health-related scoring system is difficult to evaluate with accuracy, being part science and to a large extent an art form [13]. However, it may be evaluated by assessing the ability of the scoring system to distinguish between groups of patients with expected differences in disease severity and by showing that correlation with other accepted measures of disease severity are in the direction and magnitude that would be expected if the instrument under the study is working the way it should [20].

HUANG *et al.* [14] suggested that colonization with *S. aureus* alone in adult CF patients was more likely to be found in patients with mild disease. The onset of airway colonization with *P. aeruginosa* has been regarded as a turning point that is associated with a poorer prognosis [15, 16]. The persistent recovery of mucoid *P. aeruginosa* was associated with decreased pulmonary function, poorer prognosis and an increased rate of decline in pulmonary function [16]. Furthermore, MOUTON *et al.* [17] demonstrated a strong association between increased resistance to antipseudomonal antibiotics and both the frequency of hospital admissions and the number of days spent in hospital in 34 chronically colonized CF patients [17]. In the present study, as shown in table 2, the total score and all the subscores distinguished between those patients with milder disease severity (*S. aureus* colonization) and those patients with more severe (*P. aeruginosa* colonization, mucoid or nonmucoid) as well as the worst disease severity (multiple resistant pseudomonads). There were no patients in the present study with *B. cepacia* colonization.

Several studies in the paediatric CF population have shown a good correlation between a number of radiological scores and pulmonary function testing [21–23]. In a study of 27 adult CF patients, ROSENBERG *et al.* [24] found a good correlation between the Brasfield radiological scoring system and FEV_1 % pred ($r=0.68$, $p<0.001$). The present study confirmed this good correlation, suggesting the construct validity of the Huang minor modification of the Brasfield radiological scoring system. However, as suggested by SAWYER *et al.* [23] these correlations although significant, indicated that a large proportion of the variability in radiological scores could not be explained by lung function measurements. Several studies have also shown a good correlation between clinical indices of disease severity in CF and tests of pulmonary function [25–28]. In the present study, a highly significant curvilinear

relationship between FEV_1 % pred and clinical as well as complications subscores was demonstrated. At low levels of lung function, clinical and complications subscores declined more rapidly as the FEV_1 % pred decreased. This pattern of relationship could have significant implications in assessing therapeutic modalities when FEV_1 % pred is used as outcome variable. Small differences in FEV_1 % pred, while on the steep portion of the curve could be reflected in relatively larger differences in the corresponding score.

The validity of a scoring system as an evaluative instrument measures changes over time. This can be demonstrated by showing that changes in the instrument being investigated correlate with changes in other related measures in the theoretically derived predicted direction and magnitude [9].

There are no longitudinal studies of a CF scoring system over a prolonged period of time. Wielinski *et al.* [29], reported limited and preliminary data suggesting the usefulness of the National Institute of Health (NIH) score in assessing the progression of advanced disease in cystic fibrosis patients. In the present study, significant associations were found between changes in nonpulmonary function components of the total score and changes in spirometry overall and at different levels of disease severity over a mean period of 779±204 days, providing support for the longitudinal validity of the scoring system as an evaluative instrument.

Some of the correlations may seem disappointing. Explanations for poor correlations may include small changes in pulmonary function or a true underlying weak correlation between the two variables [30–32]. Inferences from all these correlations would have been strengthened if an *a priori* prediction had been made regarding their magnitude [8]. Small changes in disease severity as measured by spirometry may also increase the influence of repeatability and measurement error on the correlations between changes in the measured variables over the 2-yr period, thereby reducing the correlation and the measured slope of the relationship [33]. In addition, the relatively narrow range of changes observed in the patients′ spirometric values overall and particularly in the separate subgroups, increased the tendency for the data points to be clustered around the midpoint of the regressions.

As shown in table 4, there was a progressive increase in the contribution of the changes in clinical and complications subscores to the changes in total score with increasing disease severity. The changes in the radiological subscores, within the limitations of the Brasfield scoring system and within the confines of the follow-up period, were most important in mild disease. The changes in pulmonary function subscore were relatively important at all levels of disease severity. In mild disease (total score ≥ 60), changes in $FEF_{50\%}$ and $FEF_{25-75\%}$ scores predominated; in moderate to severe disease (total score ≥ 35 and <60), changes in FEV_1, FVC, as well as $FEF_{50\%}$ and $FEF_{25-75\%}$ scores predominated, whereas in advanced disease (total score <35), changes in both FEV_1 and FVC scores were important, but the magnitude of change in FVC was greater than the change in FEV_1.

In conclusion, significant evidence for the construct validity and longitudinal validity of the scoring system was demonstrated.

The total score and all the subscores distinguished between patients with expected differences in disease severity. The relationship between disease severity assessed by FEV_1 % pred and the nonpulmonary function subscores was highly significant and curvilinear (for the clinical and complications subscores) and linear (for the radiological subscore). Changes in nonpulmonary function component of total score significantly varied over time in accord with changes in spirometry, suggesting the longitudinal validity of these subscores. There was a progressive increase in the contribution of the changes in clinical subscore and complications subscore to the changes in total score with increasing disease severity.

Further studies are needed to assess the responsiveness and interpretability of the scoring system in the setting of therapeutic interventions in clinical trials.

References

1. Boat TF, Welsh MJ, Beaudet AL. Cystic fibrosis. *In:* Scriver CR, Beaudet AL, Sly WS, Valle D, eds. The Metabolic Basis of Inherited Disease. New York, McGraw-Hill, 1989; pp. 2649–2680.
2. Marshall BC. Pathophysiology of pulmonary disease in cystic fibrosis. *In:* Petty TL, Cherniack RM, Fiel SB, eds. Seminars in Respiratory and Critical Care Medicine: Vol. 15; Cystic Fibrosis. New York and Stuttgart, Thieme Medical Publishers, 1994; pp. 364–374.
3. Burke W, Aitken ML. The ageing cystic fibrosis patient: presentations and problems. *In:* Petty TL, Cherniack RM, Fiel SB, eds. Seminars in Respiratory and Critical Care Medicine: Vol. 15; Cystic Fibrosis. New York and Stuttgart, Thieme Medical Publishers, 1994; pp. 383–390.
4. FitzSimmons SC. The changing epidemiology of cystic fibrosis. *Pediatr* 1993; 122: 1–9.
5. Collins FS. Cystic fibrosis: molecular biology and therapeutic implications. *Science* 1992; 256: 774–779.
6. Huang N, Helen HK, Palmer J, Hsuan F. A scoring system for short-term evaluation of patients with cystsic fibrosis: a possible means for assessment of antibiotic efficacy. *In:* Warwick WJ, ed. 1,000 years of Cystic Fibrosis Collected Papers, Minnesota University of Minnesota Dept of Pediatrics Medical School in cooperation with International Cystic Fibrosis Association National Heart Lung and Blood Institute and Fogarty International Center, 1981; pp. 207–215.
7. Matouk E, Ghezzo RH, Gruber J, Hidvegi R, Gray-Donald K. Internal consistency reliability and predictive validity of a modified N. Huang clinical scoring system in adult cystic fibrosis patients. *Eur Respir J* 1997; 10: 2004–2013.
8. Lacasse Y, Wong E, Guyatt G. A systematic overview of the measurement properties of the chronic respiratory questionnaire. *Can Respir J* 1997; 4: 131–139.
9. Guyatt G, Feeny DH. Measuring health-related quality of life. *Ann Intern Med* 1993; 118: 622–629.
10. American Thoracic Society. Standardization of spirometry 1987 update. *Am Rev Respir Dis* 1987; 136: 1285–1295.
11. 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.
12. Wagener JS, Taussig LM, Burrows B, Hernried L, Boat T. Comparison of lung function and survival patterns between cystic fibrosis and emphysema or chronic bronchitis

patients. *In:* Sturgess JM, ed. Perspectives in Cystic Fibrosis. Proceedings of the 8th International Cystic Fibrosis Congress. Mississauga, Ontario, Imperial Press, 1980; pp. 236–245.
13. McDowell I, Newell C. Measuring Health: A Guide to Rating Scales and Questionnaires. New York, NY, Oxford University Press, 1987.
14. Huang NN, Schidlow D, Szatrowski T, *et al*. Clinical features, survival rate and prognostic factors in young adults with cystic fibrosis. *Am J Med* 1987; 82: 871–879.
15. Kerem E, Corey M, Gold R, Levision H. Pulmonary function and clinical course in patients with cystic fibrosis after pulmonary colonization with *Pseudomonas aeruginosa*. *J Pediatr* 1990; 116: 714–719.
16. Demko CA, Byard PJ, Davis PB. Gender differences in cystic fibrosis: *Pseudomonas aeruginosa* infection. *J Clin Epidemiol* 1995; 48: 1041–1049.
17. Mouton JW, Den-Hollander JG, Horrevorts AM. Emergence of antibiotics resistance amongst *Pseudomonas aeruginosa* isolates from patients with cystic fibrosis. *J Antimicrob Chemother* 1993; 31: 919–926.
18. Cooper P, Robertson C, Hudson I, Phelan P. Variability of pulmonary function tests of lung in cystic fibrosis. *Pediatr Pulmonol* 1990; 8: 16–22.
19. Brazier JE, Harper R, Jones NMB, *et al*. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992; 305: 160–164.
20. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chron Dis* 1985; 38: 27–36.
21. Wong EB, Regnis J, Shier RC, Bye PT, Stewart ME. The relationship between tests of lung function and three chest radiological scoring systems in patients with cystic fibrosis. *Austras Radiol* 1993; 37: 265–269.
22. Conway SP, Pond MN, Bowler I, *et al*. The chest radiograph in cystic fibrosis: a new scoring system compared with the Chrispin-Norman and Brasfield scores. *Thorax* 1994; 49: 860–862.
23. Sawyer SM, Carlin JB, DeCampo M, Bowes G. Critical evaluation of three radiograph scores in cystic fibrosis. *Thorax* 1994; 49: 866–863.
24. Rosenberg SM, Howatt WF, Grum CM. Spirometry and chest roengenographic appearance in adults with cystic fibrosis. *Chest* 1992; 101: 961–964.
25. Beier R, Renzetti AD Jr, Mitchell M, Watanabe S. Pulmonary pathophysiology in cystic fibrosis. *Am Rev Respir Dis* 1966; 94: 430–440.
26. Matthews IW, Doershuk CF. Measurement of pulmonary function in cystic fibrosis. *In:* Rossi E, Stoll E, eds. Modern Problems in Pediatrics. Vol 10. New York, S Krager AG, 1967; p. 237.
27. Godfrey S, Mearns M. Pulmonary function and response to exercise in children with cystic fibrosis. *Arch Dis Child* 1971; 46: 144–157.
28. Featherby FA, Weng T-R, Crozier DN, Duic A, Reilly BJ, Levison H. Dynamic and static lung volumes, blood gas tensions and diffusing capacity in patients with cystic fibrosis. *Am Rev Respir Dis* 1970; 102: 737–749.
29. Wielinski CL, Warwick WJ, Budd JR. Mortality and progression of the NIH clinical and prognostic score. *Pediatric Pulmonol* 1990; 16: Suppl. 5, 259–260. (Abstract).
30. Guyatt GH, Thompson PJ, Berman LB, *et al*. How should we measure function in patients with chronic heart and lung disease? *J Chron Dis* 1985; 38: 517–524.
31. Wegener RE, Jorres RA, Kirsten DK, Magnussen H. Factor analysis of exercise capacity, dyspnoea ratings and lung function in patients with severe COPD. *Eur Respir J* 1994; 7: 725–729.
32. Wijkistra PJ, Ten Vergert EM, van der Mark TW, *et al*. Relation of lung function, maximal inspiratory pressure, dyspnea and quality of life with exercise capacity in paients with chronic obstructive pulmonary disease. *Thorax* 1994; 49: 468–472.
33. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. *Am Rev Respir Dis* 1992; 145: 1321–1327.