Internal consistency reliability and predictive validity of a modified N. Huang clinical scoring system in adult cystic fibrosis patients

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

Cystic fibrosis (CF) is a complex, multisystem inherited disorder affecting children, many of whom now live into adulthood. Pulmonary complications account for over 90% of the morbidity and mortality associated with the disease [1]. A comprehensive scoring system could be used: 1) to provide a composite index of disease status for the patient at a point in time; 2) to follow progression of the disease and predict outcome; 3) to assess outcome of therapeutic interventions in clinical trials; 4) to assist in decisions regarding need for more aggressive therapy, particularly lung transplantation; and 5) to prioritize patients awaiting lung transplantation.

Several clinical scoring systems for CF patients have been developed over the past 40 yrs [2–4]. The most widely used method of evaluation was introduced by SHWACHMAN and KULCZYCKI [2] in 1958. This system continues to be popular for classifying patients into various categories of disease severity in longitudinal studies. However, it does not include pulmonary function testing, scoring for hypoxaemic and or hypercapnic respiratory failure, or other serious pulmonary complications that influence outcome.

In 1973, TAUSSIG et al. [3], introduced a more comprehensive scoring system (the National Institute of Health) NIH scoring system. However, the criteria for scoring several items were poorly defined. These were composed of several elements or had point ranges for scoring that were not clearly graduated [5]. Furthermore, scoring for hypoxaemic and or hypercapnic respiratory failure was not included.

In 1981, HUANG et al. [6], described the use of a scoring system for the short-term evaluation of patients with CF as a means for assessing the efficacy of antibiotic treatment for acute exacerbation. A high degree of inter-rater reliability and consistency was noted in most of its components both pre- and post-treatment. More recently, SOCKRIDER et al. [5] described the evaluation of a modified NIH scoring system: it did not include scoring for respiratory failure; the general sub-score had a low internal consistency; and most of the variability in the modified NIH score was attributable to the pulmonary sub-score. The analyses suggested a new scoring structure for the NIH score. We have developed a scoring system for adults modified from the one proposed by HUANG et al. [6]. We wished to evaluate the
modified scoring by assessing its internal consistency reliability as well as its utility in predicting mortality in a population of adults with CF. 

Methods

Scoring system

We chose the N. Huang scoring system as the basis to build a comprehensive scoring for adult CF patients, because it encompassed all the pulmonary manifestations of CF, including respiratory failure. Furthermore, the scoring structure was well delineated and criteria for scoring all the items were well defined. The original general structure was maintained, but several modifications were made in order to adapt the system to current standards of measurement and to underscore applicability and relevance to the adult CF population. The modifications are shown in Appendix and discussed below.

Clinical subscore. Fever and leucocytosis item scores were removed from the original score because they rarely occurred in the absence of a pulmonary exacerbation. Weight item score was added and modified: we used the 1983 Metropolitan Life Insurance reference weights for heights to calculate the per cent predicted ideal body weight according to gender, stature and height [7]. We used the classification for nutritional status proposed by the consensus committee of the American cystic fibrosis foundation which is a modification of that by WATERLOW and RUTISHAUSER [8, 9].

Weight change item score was added and modified: weight loss was quantified as a percentage of body weight lost. Most authors consider a 10% loss of body weight occurring during the present illness to be clinically significant. Furthermore, a history of unintentional weight loss of 10% or greater over a 6 month period can be indicative of protein energy malnutrition [10, 11]. The percentage of weight loss was calculated from the last visit compared to the best recorded over the past 12 months. If the patient stabilized, albeit at a lower weight, over the past 12 months, a score of 5 points was given to emphasize his or her stability.

Dyspnoea item score replaced the original physical activity item score: we used the British Medical Research Council (MRC) Scale, modified by the American Thoracic Society [12]. Sputum item score was added: the colour of the sputum was selected as the common denominator to describe the sputum. In our experience, a darker sputum colour, in the absence of an exacerbation, suggests of the sputum was selected as the common denominator to describe the sputum. In our experience, a darker sputum colour, in the absence of an exacerbation, suggests a more advanced disease severity. Physical examination item score was also modified, as shown in Appendix.

Respiratory rate item score was modified and included the respiratory rate, breathing pattern of respiratory distress as well as cardiac frequency. The latter was included based on the study by Sini et al. [13] suggesting that an increased sinus cardiac frequency was a significant predictor of death in a group of 69 patients with chronic obstructive pulmonary disease enrolled in the nocturnal oxygen therapy trial.

Sputum culture item score was modified: Stenotrophomonas maltophilia, Serratia marsceccens, Klebsiella ozae-nae, Escherichia coli, Acinetobacter spp. and Alcaligenes spp. were arbitrarily included in the item score severity of mucoid Pseudomonas based on our clinical experience in some patients with predominant growth of these organisms. Further studies are needed to evaluate their impact on the course of CF. The presence of persistent multiply resistant Pseudomonas in respiratory cultures (defined as resistance to all the antibiotics included in all the different classes on at least two samples of sputum close to the scoring visit) was included in the worst item score severity.

Pulmonary function subscore. Forced expiratory flow at 75% of forced vital capacity (FEF75%) was deleted. Forced expiratory volume in one second (FEV1) and forced expiratory flow at 50% of forced vital capacity (FEF50%) item scores were added with the option of replacing the latter with residual volume (RV)/total lung capacity (TLC) ratio, if available. In the present study, RV/TLC ratio was not available on all patients, so FEF50% was used. Further studies using RV/TLC ratio could elucidate its role in the scoring system. With the exception of the optional item RV/TLC ratio, the specific criteria for scoring the other pulmonary function item scores were modified. In quantifying impaired forced vital capacity (FVC)% predicted, we used the Morris five-point scale [14]. In quantifying impaired FEV1 % pred, we used a modification of the Morris five point scale [14]. We used the quantification of impaired FEV1/FVC ratio derived from MILLER [15] for patients aged ≤39 yrs. For both forced midexpiratory flow (FEF25–75%) and FEF50%, we chose the lower limits of normal, close to 60% pred, which is a slightly higher value than the one suggested by Knudson et al. [16] and Paolelli et al. [17], and we used a modification of the five-point scale suggested by Morris [14] to quantify the degree of impairment.

Complications score. Haematemesis was removed. Pulmonary exacerbation and pulmonary surgery item scores were added (modified from the NIH scoring system). 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 cephalaxin; or inhaled tobramycin + p.o. antibiotics). A five-point deduction was allowed if the patient had an exacerbation now or over the past 3 months. A two-point deduction was allowed if the patient had an exacerbation over the past 12 months. A maximum deduction of seven points was allowed.

The respiratory failure item score was modified: both hypoxaemia and hypercapnia influence prognosis in CF [18, 19]. An arterial oxygen tension (PaO2) of 8.0 kPa (60 mmHg) was chosen as the demarcation between hypoxaemia and nonhypoxaemia because it is the most commonly used cut-off for clinical prescription of supplemental oxygen [20]. We stratified the severity of respirator failure as shown in Appendix.

Patient population

One hundred and nine patients were followed at the Montreal Chest Institute (MCI) between 1967 and 1994. Ninety nine had been transferred from the Montreal Children's Hospital and ten were diagnosed for the first time after the age of 18 yrs. All patients had a positive sweat chloride test.
Data collection

The modified scoring system was first developed in 1991 and used prospectively on 59 patients during a period of follow-up between December 1991 and July 1994 (31 months). Retrospective data used in the scoring were abstracted on 50 patients seen between 1967 and November 1991, through review of chest radiographs and medical records. Data from visits of patients in acute exacerbations were not included in the analyses.

In the prospective group, subjective data (cough, sputum, dyspnea and appetite) were obtained from each patient at each visit using a one-page multi-item questionnaire completed by the patient. Objective data obtained at each visit included: respiratory frequency, cardiac frequency, temperature, height, weight and physical examination. Spirometry (FEV1 % pred, FVC % pred, FEV1/FVC ratio, FEF25-75 % pred, FEF50% % pred) was performed according to standards established by the American Thoracic Society [21]. The best of two reproducible efforts performed was recorded. With the exception of FEV1/FVC ratio, they were each expressed as a percentage of the predicted normal value for the patient’s age, gender, and height (the standards were adapted from those of KNUDSON et al. [16]). Oxygen saturation was obtained at each visit by fingertip pulse oximeter (Nellcor N-100; Nellcor Puritan Bennett, Pleasanton, CA, USA). Arterial blood gases were obtained if oxygen saturation was <90% and/or when clinically indicated (particularly those of KNUDSON et al. [16]). Oxygen saturation was <90% and/or when clinically indicated (particularly in patients with FEV1 % pred <34%).

Preclinical blood gases were obtained at each visit by fingertip pulse oximeter (Nellcor N-100; Nellcor Puritan Bennett, Pleasanton, CA, USA). Arterial blood gases were obtained if oxygen saturation was <90% and/or when clinically indicated (particularly in patients with FEV1 % pred <34%). Previous studies by WAGENER et al. [19] have shown that it was rare for a patient with a FEV1 ≥34% pred to have a high arterial carbon dioxide tension (Paco2), and 47.9% of the patients with CF with an FEV1 <34% pred were CO2 retainers.

Chest radiographs were scored by our staff radiologist (RH), who had no knowledge of the other components of the score. The chest radiographs used in the scoring at the last recorded visit were taken either at the time of the visit itself, or 1–2 months from that visit, while the patient was considered clinically stable.

In the retrospective group, the same data were readily obtained from out-patient visits or hospitalizations. The subjective data (cough, sputum, dyspnea and appetite) were part of the routine history documented in the medical records of the patients.

Patients who received transplants were included in the cross-sectional analysis. The date of scoring was at the last visit prior to the transplant. Those patients who died post-transplant (n=3) were excluded from the number of deaths used in the probability of death analysis.

A relational database program (available upon request) was used for data entry and scoring.

Statistical analysis

We first evaluated the scoring system for reliability in the 109 patients by measuring its internal consistency as well as the correlations of the individual item scores and subscores to the total score, after correction for overlap. Cronbach coefficient alpha was used as a measure of internal consistency [22–24]. It was estimated for the total score with only the individual item scores included and with only the subscores included, as well as when each individual item score or individual subscore was excluded, and for each subscore separately. Although there is no fixed level of correlation that indicates reliability [24], HELMSTADTER [23] suggested that reliability is considered acceptable when alpha ≥0.5. More recently, MCDOWELL and NEWELL [24] recommended an internal consistency coefficient of ≥0.85 as acceptable.

The correlation between each individual item score or subscore and total score computed from the other item scores or subscores was assessed by Pearson correlation coefficient. Significance was accepted at p≤0.05.

We then assessed the predictive validity by estimating the probability of death at different levels of total score, FEV1 % pred and FVC % pred. The probability of death analysis was performed with the observed visits grouped by total score, FEV1 % pred or FVC % pred according to case, and then extracting the oldest observation for each subject. Kaplan-Meier survival curve was computed in the selected observations.

We then estimated the crude relative risk of death at 6, 12, 18 and 24 months using total score and FEV1 % pred. The points chosen corresponded to the approximate optimal points from the respective receiver operating curve (ROC). Fourteen patients who were lost to follow-up were subtracted from the total of 109 patients, as well as the four patients who underwent transplantation. This was done because the last recorded observation was more than 5 yrs old for the 14 patients lost to follow-up and the censoring was not random in the transplant patients. In order to avoid closed repeated observations in some patients, we filtered out all the observations with a window of 1 month, resulting in 415 observations on 91 patients at 6 months, 297 observations on 88 patients at 12 months, 189 observations on 64 patients at 18 months and 168 observations on 44 patients at 24 months. To each observation, we attached the survival status/censor at 6, 12, 18 and 24 months.

We also used “AVAS” procedure in “S-Plus 3.3” (MathSoft Inc., Seattle, WA, USA) in the reduced data set to model the contribution of FEV1 % pred, total score and subscores to the survival of the patients [25]. Initially, we used total score and FEV1 % pred; we then used all the subscores and FEV1 % pred. Survival status was coded separately in five variables, according to whether death occurred at 6, 12, 18, 24 or 36 months.

Results

Patient population and visits

Table 1 shows the characteristics of the 109 patients in the study at the time of the last recorded visit. The average duration of follow-up since the patients were first seen at the Montreal Chest Institute (MCI) was 8±5 yrs (range 7 months to 21 yrs). In the 50 patients of the retrospective group: 14 patients were alive at the time of the last recorded visit, and were subsequently lost to follow-up; two patients left the clinic and were known to be alive at the end of the study; six patients left the clinic and died, the date of death confirmed in all; and 28 patients died during follow-up at the MCI, two following transplantation. The average duration of follow-up since first seen at the MCI was 6±5 yrs. The scoring was performed on an average of four visits per patient
Values are presented as mean±SD and range in parentheses, or for sex as number of patients. FEV1: forced expiratory volume in one second; % pred: percentage of normal predicted value; FVC: forced vital capacity; IBW: ideal body weight.

during the follow-up period since the patient was first seen at the MCI.

Of the 59 patients in the prospective group, 51 were alive and still being followed. One patient was alive 4.5 yrs after lung transplantation, and two patients left the clinic and were alive at the last visit. Five patients died during the study period, of whom one died 12 months after lung transplantation. The average duration of follow-up since first seen at the MCI was 8±5 yrs. The scoring was performed at every visit during the follow-up period 1991–1994, giving an average of 19 visits per patient, and an average of four visits per patient prior to this period.

**Internal consistency reliability of the scoring system**

Table 2 shows the Cronbach alpha and correlation coefficients between all individual items and total score. The estimated Cronbach coefficient alpha was 0.934, suggesting that 93% of the observed variance in total score was consistent variance across the individual items of the scoring system. When only the subscores were included (table 3), the estimated Cronbach coefficient alpha was 0.887 for the combined total group of patients, and for the prospective and retrospective groups separately. When the clinical subscore, the pulmonary function subscore and the complications subscore were excluded one at a time, the Cronbach coefficient alpha dropped to 0.811, 0.837 and 0.843 respectively, underscoring the importance of these subscores in contributing to the reliability and accounting for the variance of the total score. Deleting the radiographic subscore resulted in little change in the coefficient alpha. The Cronbach coefficient alpha for the clinical, radiographic, pulmonary function and complications subscores were 0.914, 0.844, 0.940 and 0.606, respectively, suggesting significant internal consistency and reliability of the corresponding subscores.

With the exception of parenchymal consolidation and pulmonary surgery item scores, all the other item scores and subscores showed a significant correlation with total score after correction for overlap (tables 2 and 3). These two items were infrequent in our cystic fibrosis patients and showed little variability.

<table>
<thead>
<tr>
<th>Deleted item</th>
<th>Mean</th>
<th>sd</th>
<th>Range</th>
<th>r</th>
<th>Cronbach Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td>3.4</td>
<td>1.6</td>
<td>1–5</td>
<td>0.623</td>
<td>0.931</td>
</tr>
<tr>
<td>Weight change</td>
<td>4.0</td>
<td>1.3</td>
<td>1–5</td>
<td>0.337</td>
<td>0.934</td>
</tr>
<tr>
<td>Dyspnoea</td>
<td>3.2</td>
<td>1.3</td>
<td>1–5</td>
<td>0.875</td>
<td>0.928</td>
</tr>
<tr>
<td>Cough</td>
<td>2.5</td>
<td>0.9</td>
<td>1–5</td>
<td>0.728</td>
<td>0.931</td>
</tr>
<tr>
<td>Sputum</td>
<td>2.3</td>
<td>1.1</td>
<td>1–5</td>
<td>0.765</td>
<td>0.930</td>
</tr>
<tr>
<td>Physical examination</td>
<td>2.7</td>
<td>1.6</td>
<td>0.5–20</td>
<td>0.879</td>
<td>0.927</td>
</tr>
</tbody>
</table>

**Table 3. – Descriptive statistics, Cronbach alpha and correlation coefficients for the subscores**

<table>
<thead>
<tr>
<th>Deleted subscore</th>
<th>Score</th>
<th>r</th>
<th>Cronbach alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical</td>
<td>29.8±9.3</td>
<td>0.889</td>
<td>0.811</td>
</tr>
<tr>
<td>(10–48)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiological</td>
<td>14.9±3.5</td>
<td>0.771</td>
<td>0.900</td>
</tr>
<tr>
<td>(8–23)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulmonary function</td>
<td>12.3±6.6</td>
<td>0.809</td>
<td>0.837</td>
</tr>
<tr>
<td>(5–25)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complications</td>
<td>-10.4±8.3</td>
<td>0.795</td>
<td>0.843</td>
</tr>
<tr>
<td>(0–28)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Values are mean±SD, and range in parentheses. One hundred and nine patients were included in the model, with a Cronbach alpha of 0.887. r: correlation coefficient between each subscore and total score computed from the other subscores.
**Predictive validity**

Predictive validity was assessed by computing probability of death curves at different levels of FVC % pred FEV1 % pred, and total score (fig. 1). For patients with an FEV1 ≤28% pred, FVC ≤47% pred and total score ≤45, the probability of death at 24 months was 50%. The spirometric values closely corresponded to the previously reported values by Kerem et al. [18] and Graesemann et al. [26].

We also estimated the crude relative risk of death at 6, 12, 18 and 24 months using total score and FEV1 % pred. As shown in table 4, there was a progressive

### Table 4. – Relative risk of death at 6, 12, 18 and 24 months using total score and forced expiratory volume in one second (FEV1) % predicted

<table>
<thead>
<tr>
<th></th>
<th>Total score*</th>
<th>FEV1 % pred*</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 months</td>
<td>≤30</td>
<td>&gt;30</td>
</tr>
<tr>
<td>Dead n</td>
<td>37 21</td>
<td>17 340</td>
</tr>
<tr>
<td>Alive n</td>
<td>38 23</td>
<td>33 324</td>
</tr>
<tr>
<td>Relative risk</td>
<td>13.4 (4.10–22.15)</td>
<td>7.09 (4.81–10.32)</td>
</tr>
<tr>
<td>12 months</td>
<td>≤35</td>
<td>&gt;35</td>
</tr>
<tr>
<td>Dead n</td>
<td>56 23</td>
<td>16 202</td>
</tr>
<tr>
<td>Alive n</td>
<td>22 23</td>
<td>25 193</td>
</tr>
<tr>
<td>Relative risk</td>
<td>9.66 (5.90–15.80)</td>
<td>6.18 (4.16–9.18)</td>
</tr>
<tr>
<td>18 months</td>
<td>≤38</td>
<td>&gt;38</td>
</tr>
<tr>
<td>Dead n</td>
<td>68 20</td>
<td>16 85</td>
</tr>
<tr>
<td>Alive n</td>
<td>32 27</td>
<td>27 74</td>
</tr>
<tr>
<td>Relative risk</td>
<td>4.88 (3.07–7.75)</td>
<td>2.89 (2.05–4.07)</td>
</tr>
<tr>
<td>24 months</td>
<td>≤41</td>
<td>&gt;41</td>
</tr>
<tr>
<td>Dead n</td>
<td>75 21</td>
<td>14 58</td>
</tr>
<tr>
<td>Alive n</td>
<td>41 31</td>
<td>41 41</td>
</tr>
<tr>
<td>Relative risk</td>
<td>4.02 (2.48–6.51)</td>
<td>2.01 (1.52–2.65)</td>
</tr>
</tbody>
</table>

Values are presented as the number of observations (n), or, for relative risk, the mean and 95% confidence interval. *: the points chosen corresponded to the approximate optimal points from the respective receiver operating curve (ROC). For details of statistical analysis, see text.

---

**Fig. 1.** – Probability of death curves at 24, 18, 12 and 6 months utilizing: a) forced vital capacity (FVC) % predicted; b) forced expiratory volume in one second (FEV1) % pred; and c) total score. A: 6 months; B: 12 months; C: 18 months; D: 24 months.

**Fig. 2.** – Contribution of total score and FEV1 % predicted to the probability of survival at: a) 6; b) 12; c) 18; and d) 24 months. Survival was modelled using total score and FEV1 % pred. (Contribution to probability of survival at 36 months not shown). For definitions see legend to figure 1.
increase in the crude relative risk of death from 24 months to 6 months, suggesting increasing certitude of both total score and FEV1 % pred in predicting short-term survival. However, the crude relative risk of death values were significantly higher using total score, compared to FEV1 % pred, suggesting better prognostic discrimination of the total score.

We then used "AVAS" procedure in "S-Plus" (MathSoft Inc.) to model the contribution of the total score and FEV1 % pred to the survival of our patients at 6, 12, 18, 24 and 36 months. The corresponding r² values of the five regressions were 0.51, 0.49, 0.41, 0.39 and 0.36, respectively. The perfect progression of the r² values corroborated the increasing incertitude in predicting long-term survival in this patient population.

Both FEV1 % pred and total score were significant predictors of survival at 36 and 24 months. However, as shown in figure 2, there was a progressive decrease in the discriminating power of FEV1 % pred for predicting survival at 24, 18, 12 and 6 months, i.e. the graph turned into a vertical bar where we could see high and low estimated survival for the same value of FEV1 % pred. On the other hand, total score showed a progressive and consistent pattern that clearly predicted outcome. Figure 2 suggested that a total score of <41, <38, <35 and <30 points was associated with a poor prognosis for survival at 24, 18, 12 and 6 months, respectively. To further clarify the contribution of the different subscores and FEV1 % pred, we regressed the survival of our patients at 6, 12, 18, 24 and 36 months using all the subscores and FEV1 % pred. The corresponding R² values of the five regressions were 0.61, 0.54, 0.51, 0.43 and 0.36, respectively. As shown in figure 3, FEV1 % pred presented a similar pattern, with somewhat less discrimination for survival at 24 months. The pulmonary function and radiological subscores showed a similar but less discriminating and more noisy pattern. The complications subscore showed a progressive and consistent pattern; and the clinical subscore paralleled the total score in clearly predicting the outcome. A clinical subscore <26 points or a complications subscore >8 points was associated with a poor prognosis for survival at 36 and 24 months. Furthermore, a clinical subscore <23 points was associated with a poor prognosis for survival at 18 and 12 months; and clinical subscore <20 points was associated with a poor prognosis for survival at 6 months.

Discussion

We described a modified N. Huang scoring system and its application in a population of adult CF patients. The necessary data could be readily abstracted from the patient's visit to a CF centre. They were part of the routine history and physical examination, and included laboratory data such as microbiology, radiology, spirometry, oximetry and arterial blood gases measurements when necessary. A relational data base program allowed easy data entry (about 2–3 min), instantaneous scoring and longitudinal follow-up data on all of the patients. In the absence of a computerized system, and after all the data had been collected, it took about 4–5 min to compute the score by hand.

Although we used $P_aO_2$ to assess hypoxaemia in our patients, further studies are needed to assess the gradation of hypoxaemia using $O_2$ saturation by pulse oximetry instead of $P_aO_2$, and identify the items that increase the likelihood of detecting CO₂ retention. In general, a $P_aO_2$ of 7.3 kPa (55 mmHg) corresponds to a saturation

![Fig. 3. Contribution of clinical subscore, complications subscore and FEV1 % predicted to the probability of survival at: a) 6; b) 12; c) 18; and d) 24 months. Survival was modelled using all the subscores and FEV1 % pred. (Contribution of radiological and pulmonary function subscores not shown). For definitions see legend to figure 1.](image-url)
of 88% [20]. In a study of 55 patients with chronic lung
disease and hypoxaemia, CARLIN et al. [27] found that
more than 80% of patients with PaO, less than 55 mmHg
had cutaneous oximetry greater than 85%. A proposed
gradation would be: O2 saturation 88–89%: four-point
deduction; O2 saturation 85–87%: six-point deduction;
O2 saturation ≤85%: eight-point deduction. This nonin-
vasive method would be of particular value in those pati-
ents refusing an arterial puncture.

The scoring system did not directly address some of
the nonpulmonary complications encountered in this pop-
ulation, namely: liver cirrhosis; distal intestinal obstruc-
tion syndrome (DIOS); gall-bladder disease; and diabetes mellitus, unless these complications affected other items
in the scoring. Thus, although liver cirrhosis accounted
for 1–2% of CF mortality each year, the main problem in
evaluating the contribution of liver disease to mor-
bidity and mortality in CF was the lack of suitable vari-
able for prognostic analysis [28, 29].

We first evaluated the scoring system for internal con-
sistency reliability. Ideally, items within a scoring index
would be closely correlated and have a high internal
consistency and a coefficient alpha close to 1.0 [6]. The estimated Cronbach coefficient alpha for the total score,
clinical, pulmonary function and radiological subscores
were high, indicating a high degree of internal consisten-
cy of their individual items in measuring the correspond-
ing score. The modest value of the Cronbach coefficient
alpha for the pulmonary complications subscore proba-
bly reflected the infrequent occurrence of some of these
complications, some predominantly in advanced stages
of the disease and showing little variability.

The validity of the present scoring system as a mea-
sure of a patient’s health status could not be determined
by comparing it to a gold standard. No true gold stan-
dard exists. Since outcomes such as death are clear cut,
we validated the scoring system as a predictor of death
(predictive validity). As shown in figure 1, the curves
computed with the total score were steeper than the curves using FEV1 % pred or FVC % pred, suggesting
better discrimination in the last 2 yrs of life. Such an
improved prognostic discrimination was further corro-
borated by the significantly higher crude relative risk
of death at 24, 18, 12 and 6 months, using the total
score compared to FEV1 % pred.

Although both total score and FEV1 % pred were signif-
ificant predictors of survival at 36 and 24 months, the
prognostic value of the total score became progressive-
ly more discriminating, compared to FEV1 % pred, in
predicting survival at 24, 18, 12 and 6 months (fig. 2).
Progressive and consistent changes were recorded in
total score, predominantly through deteriorating clin-
ical and complications subscores (fig. 3), whereas radi-
ological and pulmonary function subscores, as well as
FEV1 % pred, tend to lose discriminating power. These
findings were not surprising for the radiological scores.
With the widespread changes usually present in the lungs
of adult CF patients, radiological scores may be less
sensitive in assessing pulmonary disease status at a point
in time and in demonstrating significant changes over
time [30]. This is particularly true in patients with end-
stage lung disease as shown in our study.

The results of the present study, as well as others,
confirm that an FEV1 % pred <30% was a significant
predictor of survival [18, 26]. However, GRASEMANN et
al. [26] demonstrated a high degree of interindividual
variability in survival of CF patients with an FEV1 %
pred <30% (mean ±sd 3.2±3.2 yrs, range 0–8 yrs). It
was therefore suggested that the decision for lung trans-
plantation should not be based on pulmonary function
alone. Previous studies used separate parameters such
as pulmonary function, blood gas levels and nutrition-
al status in the evaluation of CF patients on the wait-
ing list for lung transplantation [31, 32].

In the present study, the total score added a further
dimension to these parameters; they were viewed in the
context of the overall clinical picture. Our results sug-
gested that the total score offered a better discrimina-
ting scale, compared to FEV1 % pred, in the prognostic
evaluation of the end-stage CF patients. It could be of
value in decisions regarding referral of CF patients for
lung transplantation, prioritizing patients awaiting lung
transplantation and more reliably identifying those pati-
ents at a higher risk of early death so as to intensify
their medical therapy and improve their chances of sur-
viving to transplantation.

Further prospective studies are needed to confirm the
prognostic value of the proposed scoring system and to
evaluate its relationship to other outcome variables relat-
ed to prognosis in CF patients, such as aerobic fitness
and the 12 min walking distance [33, 34]. Furthermore,
the scoring system cannot be considered fully validat-
ed until it has been applied independently to a differ-
ent population of adults with CF.

In conclusion, the proposed modified scoring showed
a high degree of internal consistency reliability. All the
individual item scores and subscores significantly con-
tributed to the reliability and variance of the total score.
Predictive validity of the scoring system was also demon-
strated. The total score was progressively more discrim-
inating, compared to forced expiratory volume in one
second expressed as a percentage of the predicted value,
in predicting survival of cystic fibrosis patients at 24,
18, 12 and 6 months.

Appendix: proposed modified N. Huang scoring
system. Instructions for scoring.

Clinical subscore (50 points)

1. Weight

   5 = normal weight (ideal body weight)
   4 = 85–90% ideal body weight
   3 = 80–84% ideal body weight
   2 = 75–79% ideal body weight
   1 = <75% ideal body weight

2. Weight change (compared to the best previous over
   the past 12 months)

   5 = no weight loss
   4 = weight loss 2–3%
   3 = weight loss 4–5%
   2 = weight loss 6–9%
   1 = weight loss ≥10%
3. Dyspnoea
5 = dyspnoea 0; not troubled with breathlessness except with strenuous exercise
4 = dyspnoea 1; troubled with shortness of breath when hurrying on the level or walking up a slight hill
3 = dyspnoea 2; walks slower than people of the same age on the level because of shortness of breath or has to stop for breath when walking at one’s pace on the level
2 = dyspnoea 3; stops for breath after walking about 100 m or after few minutes on the level
1 = dyspnoea 4; too breathless to leave the house or breathless when dressing or undressing

4. Cough
5 = no cough even with physical therapy and postural drainage
4 = cough with physical therapy and postural drainage or dry occasional cough
3 = infrequent productive cough 1–5 times per day
2 = frequent productive cough more than 5 times per day including night cough
1 = frequent severe productive cough or coughing spells once each 30–60 min

5. Sputum
5 = none most of the time
4 = clear or white
3 = yellow or beige
2 = pale green
1 = dark green or brown or dark grey

6. Physical examination (air entry/crackles/hyperinflation)
5 = normal air entry, or lungs clear, or no hyperinflation
4 = mildly reduced air entry, or lungs clear, or mild hyperinflation
3 = moderately reduced air entry, or localized inspiratory crackles and/or one area of bronchial breathing or moderate hyperinflation
2 = moderately severe reduction in air entry, or 2–3 areas of inspiratory crackles, and/or areas of bronchial breathing, or moderately severe hyperinflation
1 = severe reduction in air entry, or diffuse inspiratory crackles and/or more than areas of bronchial breathing or severe hyperinflation
The score chosen will correspond to only one or another item on the same line.

7. Respiratory rate/breathing pattern/cardiac frequency (respiratory rate combined score)
5 = normal respiratory frequency for age, or normal breathing pattern, or normal cardiac frequency
4 = slightly faster than normal respiratory frequency (12–14 breaths-min⁻¹) or no use of accessory muscles and or no intercostal retraction or normal cardiac frequency
3 = moderate increase (14–20 breaths-min⁻¹), or mild use accessory muscles and or mild intercostal retraction, or mild increase cardiac frequency 90–100 beats-min⁻¹.
2 = moderately severe tachypnoea (20–26 breaths-min⁻¹), or moderate use of accessory muscles and or moderate intercostal retraction, or moderate increase in cardiac frequency to 100–120 beats-min⁻¹.
1 = marked tachypnoea (26 breaths-min⁻¹), or severe use of accessory muscles and or severe intercostal retraction and/or interference with carrying out regular conversation and/or paradoxical abdominal movements, or severe increase cardiac frequency >120 beats-min⁻¹.
The score chosen will correspond to only one or another item on the same line.

8. Culture
5 = Normal throat flora
4 = Staphylococcus aureus and/or other gram-positive cocci with or without Haemophilus influenzae
3 = Pseudomonas aeruginosa, nonmucoid or Pseudomonas fluorescens or nonmucoid Pseudomonas spp.
2 = Pseudomonas aeruginosa, mucoid and or mucoid Pseudomonas spp., Stenotrophomonas maltophilia, Serratia marcescens, Klebsiella ozaenae, Escherichia coli, Acinetobacter spp. and Alcaligenes spp.
1 = Burkholderia cepacia and/or the presence of multiply resistant Pseudomonas aeruginosa or species nonmucoid or mucoid
In the presence of multiple organisms, the organism in the worst category will dictate the score.

9. Appetite
5 = greater than normal (two servings of food at each meal)
4 = good appetite (consuming slightly more than a regular serving of food at each meal)
3 = average appetite (consuming a regular serving of food at each meal)
2 = less than average appetite (consuming slightly less than a regular serving of food at each meal)
1 = poor appetite

10. General condition
5 = cheerful, energetic
4 = pleasant, sociable with peers, but appearing chronically ill
3 = depressed, irritable, unco-operative, uncomfortable appearance
2 = very sick and depressed
1 = bedridden
This score represented the clinician’s assessment of the patient’s general condition

Radiographic subscore (25 points)

1. Air trapping
5 = none
4 = minimal hyperinflation
3 = mild diffuse hyperinflation with increase in anteroposterior (A-P) diameter and low diaphragm
2 = moderately diffuse hyperinflation plus change in A-P diameter and low diaphragm
1 = severe hyperinflation with flat diaphragms and dark square chest, sternal bowing with or without thoracic kyphosis
2. Linear markings (single or parallel line densities (tram lines), or end on circular densities due to bronchial wall thickening (ring shadows))

5 = none
4 = minimal impression of increased perihilar markings
3 = definite thickening confined to central and two or less peripheral areas
2 = diffuse markings extending to the periphery of the lung
1 = severe, diffuse markings ("mostly white streaks")

3. Nodular cystic lesions (discrete, small, rounded densities about 0.5 cm in diameter or larger, radiolucent or radio-opaque).

5 = none
4 = few lesions
3 = multiple lesions, but localized
2 = multiple lesions diffusely scattered
1 = almost honeycomb appearance

4. Parenchymal lesions (segmental or lobar atelectasis or consolidation including pneumonia)

5 = none
4 = single, small area involved
3 = consolidation of one or more segments of a lobe
2 = consolidation including two different lobes
1 = consolidation involving more than two lobes

5. General impression (judgement of overall severity of radiographic changes)

5 = normal appearing
4 = slightly abnormal involving one area or one quadrant
3 = mildly, but diffusely, abnormal
2 = moderately severe change
1 = severe changes present (including pneumothorax, congestive heart failure)

Pulmonary function subscore (25 points)

<table>
<thead>
<tr>
<th>Score</th>
<th>FVC (% pred)</th>
<th>FEV1 (% pred)</th>
<th>FEV1/FVC</th>
<th>FEF25-75 (% pred)</th>
<th>FEF50 (% pred)</th>
<th>RV/TLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>≥80%</td>
<td>≥80%</td>
<td>≥75%</td>
<td>≥60%</td>
<td>≥60%</td>
<td>&lt;25%</td>
</tr>
<tr>
<td>4</td>
<td>65–80%</td>
<td>60–80%</td>
<td>65–75%</td>
<td>45–60%</td>
<td>45–60%</td>
<td>25–29%</td>
</tr>
<tr>
<td>3</td>
<td>50–64%</td>
<td>50–60%</td>
<td>55–64%</td>
<td>30–44%</td>
<td>30–44%</td>
<td>30–39%</td>
</tr>
<tr>
<td>2</td>
<td>35–49%</td>
<td>35–49%</td>
<td>45–54%</td>
<td>20–29%</td>
<td>20–29%</td>
<td>40–49%</td>
</tr>
<tr>
<td>1</td>
<td>&lt;35%</td>
<td>&lt;20%</td>
<td>&lt;45%</td>
<td>&lt;20%</td>
<td>&lt;20%</td>
<td>≥50%</td>
</tr>
</tbody>
</table>

Complications subscore (maximum = 37 points)

Subtract points from total score for each serious complication

1. Pulmonary exacerbation requiring therapy max 7
   a) Present or past 3 months 5
   b) Past year 2

2. Pneumothorax max 5
   a) Past 6 months or recurrent (i.e. ≥2 episodes over any period of time) 5
   b) Ever 2

3. Haemoptysis max 5
   a) Massive past 6 months 5
   b) More than 6 months ago 2
   b) Minor in the past 3 months 1

4. Respiratory failure max 10
   a) PaO2 7.32–7.80 kPa 4
   b) PaO2 6.65–7.31 kPa 6
   c) PaO2 <7.31 kPa 8
   d) Any of above with hypercapnia 9
   e) Any above + previous intubation (in past year) 10

5. Cardiac enlargement or congestive heart failure max 5
   a) Cardiac enlargement 3
   b) Congestive heart failure 5

6. Pulmonary surgery max 10
   a) ≤one lobe 3
   b) >one lobe 5

Total score = (clinical subscore (50 points) + radiographic subscore (25 points) + pulmonary function subscore (25 points) - (complications subscore)

Acknowledgements: The authors gratefully thank R. Menzies for review of the manuscript, suggestions and comments. They also wish to thank F. Bellavance, Dept of Clinical Epidemiology St Mary’s Hospital Center, for assistance in statistical analysis. They are indebted to the entire staff of the Cystic Fibrosis Clinic for their assistance in collecting the clinical data and to S. Sabatini, Logiciel Duquesne, for assistance in the development of the relational data base program "CFK" for data entry and analysis.

References


