ORIGINAL ARTICLES

Cystic Fibrosis in Adults: Inter- and Intraobserver Agreement for the Brasfield and Chrispin-Norman Chest Radiography Scoring Systems and Correlation With Clinical Data and Spirometry

V. Gutiérrez, a M.J. Olivera, b R.M Girón, b F. Rodríguez-Salvanés, b and P. Caballero a

a Servicio de Radiodiagnóstico, Hospital de la Princesa, Universidad Autónoma, Madrid, Spain.
 b Servicio de Neumología, Hospital de la Princesa, Universidad Autónoma, Madrid, Spain.

OBJECTIVE: Most chest radiography scoring systems for patients with cystic fibrosis have been developed for children but are also used for adults. Our aim was to evaluate the intra- and interobserver variability of 2 radiographic scoring systems in adults with cystic fibrosis and to assess the correlation of these systems with clinical and spirometric parameters.

PATIENTS AND METHODS: The chest x-rays of 24 adult patients with cystic fibrosis were compared using 2 scoring systems (Brasfield and Chrispin-Norman). The x-rays were scored by 2 radiologists and reevaluated 4 months later by 1 of the 2 observers. Intra- and interobserver agreement was assessed using the intraclass and Pearson’s correlation coefficients. The radiographic scores were compared to lung function tests and other clinical data.

RESULTS: Both intra- and interobserver agreement were high (r ≥ 0.9 and the intraclass correlation coefficient ≥ 0.85 with both systems for both samples). Both scoring systems correlated with spirometry results: forced expiratory volume in the first second (FEV 1) (r = 0.64 and r = 0.55), FEV 1% (r = 0.75 and r = 0.72), and the percentage of forced vital capacity in relation to the predicted value (r = 0.63 and r = 0.056). We found no association between scoring system and sex, age, or body mass index.

CONCLUSIONS: Assessment of chest radiographs of adult patients with cystic fibrosis by the Brasfield and Chrispin-Norman scoring systems shows good intra- and interobserver agreement. Both systems correlate well with lung function variables, especially FEV 1.

Key words: Cystic fibrosis. Lung function tests. Intraobserver agreement. Interobserver agreement. Chest radiography scoring systems.

Arch Bronconeumol. 2005;41(10):553-9

O2CCONCLUSIONES: La radiografía de tórax en pacientes adultos con FQ, valorada mediante los sistemas de puntuación de Brasfield y Chrispin-Norman, presenta buenos acuerdos intra e interobservador. Ambos sistemas de puntuación presentan una buena correlación con la función pulmonar, especialmente con el FEV 1.


Introduction

Cystic fibrosis is a multisystemic inherited disease of the exocrine glands that is mainly characterized by airway infection, chronic airway obstruction, and pancreatic insufficiency. Studies have revealed that the incidence of cystic fibrosis among whites is variable,
with rates ranging from 1 in 900 (Quebec, Canada) to 1 in 5352 (Catalonia, Spain).2

Cystic fibrosis is caused by an autosomal recessive mutation of a gene located on the long arm of chromosome 7. The protein coded by the gene—the cystic fibrosis transmembrane conductance regulator (CFTR)—essentially forms a chloride ion channel and its dysfunction or inactivation generates hydroelectric changes in the glands of affected organs. The CFTR also regulates other chloride and sodium ion channels. Changes lead to a thickening of bronchial secretions, which predisposes the lung to chronic infection, and an increase in the number of neutrophils, whose lysis produces DNA and actin. Those events further increase viscosity, perpetuating the vicious cycle of obstruction, inflammation, and infection, with its consequent clinical manifestations.3

The diagnosis of cystic fibrosis is based on finding a concentration of chloride in sweat in excess of 60 mmol/L, in 2 measurements, and/or the detection of 2 cystic fibrosis mutations, and/or alteration of the nasal potential difference along with at least one of the following findings: characteristic phenotype (respiratory or digestive disorders consistent with the diagnosis or bilateral absence of the vas deferens), a family history of cystic fibrosis (siblings or cousins), and/or a positive neonatal screening test (elevated immunoreactive trypsinogen level).4 Both simple radiographs and computed tomography (CT) of the chest might suggest the diagnosis but are not specific. Images reveal hyperinflation, thickening, and mucous plugs, bronchiectasis in the upper lobes, disseminated atelectasis, and confluent infiltrates. When the disease is advanced, x-rays reveal massive hyperinflation, cysts, extensive bronchiectasis, and areas of segmental and lobar atelectasis.5

Because the last 10 years have seen the publication of several reports of bronchial wall thickening observable by high resolution CT of the lung in all of patients, of bronchiectasis in 87%, of bronchocleses in 64%, and of emphysema in 28%, thoracic CT is considered to have greater sensitivity and specificity than simple radiography.6

Several systems for assessing pulmonary lesions in the chest x-rays of patients with cystic fibrosis appeared in the 1970s, reflecting an effort to facilitate systematic description and the comparison of serial images. The first was published by Chrispin and Norman7 in 1974 and a modification was proposed by Brasfield et al8,9 5 years later. Advances in the management of cystic fibrosis have led to increased survival, such that 40% of today’s patients with this disease are older than 16 years of age, yet most evaluation systems were developed for children and have come to be applied in the same way for adults.10 Good results have been reported from a study of inter- and intrarater variability between several systems of radiographic classification,11 and some very interesting studies have shown a relation between spirometric findings and chest radiography classification in adult patients with cystic fibrosis.11,12

Our aims in the present study were to assess the inter- and intraobserver reliability of 2 chest radiography scoring systems (Brasfield scoring and Chrispin-Norman scoring), and to explore correlations between classifications by those 2 systems and the following clinical and spirometric parameters: body mass index (BMI), age, antibiotic treatment, forced expiratory volume in the first second (FEV1), forced vital capacity (FVC), and those same lung function variables expressed as percentages of a theoretical reference value (FEV1% and FVC%) in a population of adult patients with cystic fibrosis. For the interobserver reliability study, 2 independent expert observers participated.

Patients and Methods

Study Design

A cross-sectional study was carried out in the respiratory medicine and radiodiagnostic departments of Hospital Universitario de la Princesa in Madrid, Spain. Patients diagnosed with cystic fibrosis were included. All were being treated by the adult cystic fibrosis unit of the hospital and were over 18 years old, had at least 1 posteroanterior and 1 lateral chest x-ray on file, and had been treated by the unit for at least a year before the radiographs were taken. Spirometric test results and clinical status had also been entered into the records for the studied patients within 1 month of radiography. The chest x-rays included for study were from the period comprising January 2000 through September 2003 and, along with the other information gathered, had been taken as part of routine annual check-ups. At the time of the annual check-up, all patients had been stable for at least a month, with no clinical signs or symptoms of exacerbation (defined as increased expectoration or coughing or need for antibiotics). We also recorded the number of times the patient had required oral or intravenous antibiotic treatment during the year prior to the check-up. All instances of intravenous antibiotic treatment were recorded, whether it was provided in hospital, at home,13 or both. Nutritional status was recorded as the BMI (kg/m2) expressed as a percentage of the 50th percentile for healthy individuals of the same age and sex.

Patients were excluded if the aforementioned criteria were not met, if they had other pulmonary diseases such as allergic bronchopulmonary aspergillosis, or if they had received a lung transplant. The study was approved by the ethics committee of our hospital and did not require informed consent because there was no departure from normal clinical practice.

Lung Function Tests

Lung function was assessed with the Datospir 120 spirometer (Sibelmed, Barcelona, Spain). The best of 3 maneuvers was recorded on the day of the annual check-up. Both absolute values and percentages of reference values were recorded for FEV1, FVC, and the ratio of FEV1 to FVC.

Classification of Chest Radiographs

Two expert, blinded radiologists independently classified 48 chest radiographs (24 frontal and 24 lateral) for the 24 cystic fibrosis patients. Each observer was provided with a
written description of each scoring system. Neither knew the spirometric or clinical data that had been collected for each patient in the study. Four months later, the first observer reclassified 16 radiographs from the same 24 sets, following the same method.

In the Chrispin-Norman scoring system chest radiographs are assessed first according to thoracic configuration, as follows: degree of bowing of the sternum, degree of kyphosis, and degree of flattening of the diaphragm. Each parameter is scored 0 if absent, 1 if present but not marked, or 2 if present and marked. Pulmonary shadows—including lines, nodular images, rings, and large shadows—are also scored. Each hemithorax is divided into its upper and lower half, at the horizontal fissure on the right side and at the hilum on the left. The first 3 parameters (linear, nodular, or ring-shaped shadows) are scored for each of the 4 zones as 0 if not present, 1 if present but not marked, or 2 if marked. Large shadows are also scored in all 4 zones as 1 if segmental and 2 if lobar. The overall score, which is the sum of all the previous ones, is considered to indicate mild (0-8), moderate (18-27), or severe (28-38) disease. Thus, as the score increases so does radiographic severity.

Brasfield scoring takes into consideration air trapping, linear shadows (from peribronchial inflammation), nodular-cystic lesions (which indicate fusion of nodular shadows and rings in the Chrispin-Norman system and in systems that integrate microabscesses, small areas of consolidation and mucus plugs), both segmental and lobar areas of consolidation and atelectasis, and finally overall impression of severity. For air trapping, linear images, and nodular-cystic lesions, a score of 0 is given if they are absent and a score between 1 and 4 according to severity. For large lesions, a score of 0 indicates absent and a score between 3 and 5 is given according to severity. Likewise, for overall impression of severity, scores range from 0 (normal) to 5 to reflect severity. The total score is obtained by subtracting the sum from 25. Thus, a higher score reflects a better radiographic classification, with 25 indicating a chest radiograph without alterations.

Statistical Analysis
Agreement between the scoring systems was assessed using Pearson’s correlation coefficient, Bland-Altman analysis of agreement, and the intraclass correlation coefficient. Pearson’s correlation coefficient analyzes the linear correlation between 2 quantitative variables. The Bland-Altman analysis looks at the difference between the scores given by 2 observers in relation to the average of scores given by both. Thus, it provides information on the magnitude of discrepancies between each pair of observations and the presence or absence of systematic deviations between 2 determinations. The intraclass correlation coefficient estimates the average correlation between all possible orderings of pairs of observations, and therefore, corrects for error due to changes in order.

To confirm the internal validity of the study we examined the correlation between the 2 scoring systems on the one hand and the clinical and spirometric parameters on the other. To do so, we determined Pearson’s correlation coefficient between the mean of the scores given by the 2 observers on the one hand and BMI and spirometric variables on the other. We grouped patients who had received oral antibiotic treatment at least once in the year before the x-ray was taken and compared them to patients who had not been treated. We also grouped patients who received intravenous treatment during the same period so that they could be compared to those who had not. To demonstrate a statistical association between scores (quantitative variable) and the need for either oral or intravenous antibiotic treatment (dichotomized variables) we used the Student t test for parametric data given that the hypothesis of normality was not rejected by the Kolmogorov-Smirnov test.

Results
The 24 patients, 11 men and 13 women, had a mean age of 26 years (range, 20-46). The mean FEV1 in these patients was 2.335 L (69%) and mean FVC was 3.242 L (76%). All had good nutritional status and the mean BMI was 21.9 kg/m2; only 2 patients had a BMI less than 20 kg/m2 and 1 patient had a BMI greater than 25 kg/m2. The mean Brasfield score was 18.5 points; the mean Chrispin-Norman score was 9.64. These means, which were obtained by averaging the scores of the 2 observers, corresponded to moderate pulmonary involvement. Only 2 patients had Brasfield scores under 16, indicating severe pulmonary alteration, and 2 others had Brasfield scores of 24, indicating a nearly normal radiograph. Patient characteristics are shown in Table 1.

Interobserver Agreement
Pearson’s correlation coefficients were 0.901 for the Brasfield classification system and 0.911 for the Chrispin-Norman system. The intraclass correlation coefficients were 0.88 (Brasfield) and 0.85 (Chrispin-Norman) for agreement between observers.4,14 That level of agreement is considered excellent (Figures 1 and 2).

Intraobserver Agreement
Pearson’s correlation coefficients reflecting intraobserver agreement were 0.925 for Brasfield scoring and 0.939 for Chrispin-Norman scoring. The intraclass correlation

### TABLE 1
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Summary of Characteristics of the Series of Patients With Cystic Fibrosis*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex, man:woman</td>
<td>11:13</td>
</tr>
<tr>
<td>Age, years</td>
<td>26.0 (5.85) (range, 20-47)</td>
</tr>
<tr>
<td>Lung function variables</td>
<td></td>
</tr>
<tr>
<td>FEV1, %</td>
<td>65 (19.82) (range, 26-103)</td>
</tr>
<tr>
<td>FVC%,</td>
<td>76.33 (17.61) (range, 48-107)</td>
</tr>
<tr>
<td>FEV1, L/min</td>
<td>2.335 (7.57) (range, 1.130-4.480)</td>
</tr>
<tr>
<td>FVC, L</td>
<td>3.242 8 (8.81) (range, 1.090-5.240)</td>
</tr>
<tr>
<td>BMI, P50</td>
<td>22.33 (6.32) (range, 16.80-31.21)</td>
</tr>
<tr>
<td>Brasfield</td>
<td>18.5 (3.064) (range, 13.30-24)</td>
</tr>
<tr>
<td>Chrispin-Norman</td>
<td>9.64 (4.82) (range, 1-17)</td>
</tr>
</tbody>
</table>

*Data are expressed as means followed by SD and ranges within parentheses. FEV1 indicates forced expiratory volume in the first second; FVC, forced vital capacity; FEV1 % and FVC %, percentages of the reference values for the respective variables; BMI, body mass index; and P50, the 50th percentile.
The coefficient indicative of intraobserver agreement was 0.91 for both systems ($P<.0001$). The Bland-Altman analysis showed that the differences in scoring by the same observer were probably due to chance, given that the limits of agreement included 0 (Figures 3 and 4).

**Relationship Between the Scoring Systems and Clinical and Spirometric Variables**

Table 2 shows Pearson’s coefficients of correlation between the clinical variables and the average of the 2 observers’ scores by each system for classifying chest radiographs. Both systems showed good correlation with FEV$_1$, FEV$_1$%, and FVC%, and excellent correlation with FEV$_1$% was evident. We found no correlation between radiography scores and FVC or BMI.

The differences between the correlations with each of the 2 systems were not significant, although the Brasfield score tended to correlate better with spirometric data (with the exception of FVC), and the correlation with FEV$_1$, FEV$_1$%, and FVC% reached statistical significance ($P<.05$). The strongest correlation was with FEV$_1$% ($r=0.753$; $P<.001$). The Chrispin-Norman score was also significantly correlated with FEV$_1$% ($r=0.716$; $P<.001$).

**Relationship Between the Scoring Systems and Qualitative Variables**

The relationship between scores and sex, analyzed with the Student $t$ test, showed that men had lower Brasfield scores and higher Chrispin-Norman scores,
indicating that their radiographic disease severity classification was worse. The results were statistically significant for Chrispin-Norman scoring (Table 3).

The patients who required intravenous antibiotic treatment at home or in hospital at least once during the year before the chest radiograph was taken had lower Brasfield scores and higher Chrispin-Norman scores, reflecting worse radiographic severity classification. The difference was statistically significant for the Brasfield system (Table 3). Although the mean Brasfield score for patients who required oral treatment was 1 point lower than the mean for those who did not, and the mean Chrispin-Norman score was 2 points higher, those differences were not statistically significant.

Discussion

Cystic fibrosis has traditionally been considered a pediatric disease. Nevertheless, thanks to advances in our understanding and in therapy, survival has been prolonged considerably in recent decades, such that 40% of patients now live at least 16 years. We studied 24 patients, a sample size similar to that of Terheggen et al., although unlike those authors, we studied adults. Only a small portion of the literature concerns adults, the usual focus being on children.

We found that inter- and intraobserver agreement was good for both chest radiography scoring systems, especially for Brasfield scoring, when applied to adults. This indicates that the scoring systems studied are accurate and reliable and seem to be stable scales for use in this adult population. We would therefore say that the level of reproducibility of our study is high, such that the results would be similar in similar conditions. Other studies have also demonstrated low interobserver variability for chest radiography scoring, but none achieved coefficients of correlation as high as 0.911. The Bland-Altman method of analysis revealed that error was random. Limits of agreement were better for the Brasfield system. Good agreement between the 2 systems means that the classifications assigned by the 2 scorers were similar. The limits of agreement included 0 for the Brasfield system, meaning that it is possible that there were no differences between scores assigned by the 2 observers. Such was not the case for the Chrispin-Norman system, for which the confidence interval included negative values only, but not 0, meaning that scores from one of the observers were lower than those from the other within the 95% confidence interval. However, the score difference between them was 1.7 points on the average, a difference that would not have changed the classification of radiographic severity for any patient.

A slight difference was seen between the 2 scoring sessions on the intraobserver agreement analysis: scores at the second scoring session were somewhat higher on the Brasfield scale and somewhat lower on the Chrispin-Norman scale. That observation suggested to us that the error was consistent, as both variations indicated less severity. Intraobserver agreement was good for both systems, meaning that scores given at the 2 sessions were similar. Two other studies reported intraobserver coefficients of agreement over 0.919. During annual check-ups of cystic fibrosis patients, it is recommended to order a chest x-ray. Simple radiographs are useful for evaluating respiratory status, particularly for patients unable to cooperate with spirometry, such as children under the age of 5 years or those with mental illness. Recent studies indicate that high resolution CT is more sensitive than chest radiography and detects slight alterations earlier in cystic fibrosis patients, allowing a more specific diagnosis to be reached.

Although those findings are promising, high resolution CT has disadvantages in comparison with simple chest radiography. The former is more costly and involves more radiation. Some authors have
GUTIÉRREZ V, ET AL. CYSTIC FIBROSIS IN ADULTS: INTER- AND INTRAOBSERVER AGREEMENT FOR THE BRASFIELD AND CHRISPIN-NORMAN CHEST RADIOGRAPHY SCORING SYSTEMS AND CORRELATION WITH CLINICAL DATA AND SPIROMETRY

recommended low dose high resolution CT, with slices every 20 mm to reduce radiation exposure, particularly in children. In cystic fibrosis, simple chest radiography plays an important role. It allows acute complications like pneumothorax and atelectasis to be ruled out, and in association with pulmonary function tests it is a good tool for assessing progression of disease. The Brasfield scoring system, which is better correlated with lung function and easier to use, is preferable to Chrispin-Norman scoring, but both systems are very good, systematic ways to analyze the chest x-rays of cystic fibrosis patients and allow for comparisons between new and old images.

After the age of 5 years patients can perform lung function test maneuvers properly, especially if the procedure is adequately explained by trained clinicians. Spirometry, a noninvasive technique that is easy to perform and reproducible, is a good way to demonstrate the patient’s respiratory status. We found that the correlation was good between both radiography scoring systems and the lung function parameters FEV\(_1\), FEV\(_1\)% and FVC%. That correlation has been reported previously, mainly in children, and it has also been analyzed in adults. In our study FEV\(_1\)% which is one of the most important parameters for assessing disease course and predicting mortality in cystic fibrosis patients, correlated more strongly with radiographic scoring by both systems than did the other lung function variables, an observation that is also consistent with previous reports.

Cleveland et al showed that male cystic fibrosis patients had worse Brasfield scores than females over time, although clinical course and spirometric findings were worse in females. Those authors suggested that the differences could be attributed to the greater extension of pulmonary tissue in males, which in the long term would mean less loss of FEV\(_1\) and longer life expectancy in spite of greater structural damage. We also recorded worse radiography scores among the male patients on both scales.

We did not see a good correlation between age and scores on either system. Cystic fibrosis progresses as the patient ages, reflected in scores and lung function variables, an observation that is also consistent with previous reports.

Although previous publications do not assess the relation between radiography classifications and intravenous or oral antibiotic treatment received during the year prior to the x-ray, we did study that association. It would be useful to be able to predict need for intravenous or oral antibiotic treatment in function of degree of pulmonary alteration shown by chest radiography.

We conclude that there is excellent inter- and intraobserver agreement between the Brasfield and Chrispin-Norman scales when they are applied to the chest x-rays of adult cystic fibrosis patients. Brasfield scoring was the more stable of the two. Both Brasfield and Chrispin-Norman scoring of the chest x-rays of our adult cystic fibrosis patients correlated highly with lung function (FEV\(_1\), FEV\(_1\)% FVC%). Administration of antibiotics, especially intravenous infusion, correlates with lower Brasfield and higher Chrispin-Norman scores, which indicate greater radiographic severity of disease.

REFERENCES