Reliability, Repeatability, and Sensitivity of the Modified Shuttle Test in Adult Cystic Fibrosis*

Judy Bradley, DPhil; Jennifer Howard, BSc; Eric Wallace, DPhil; and Stuart Elborn, MD

Study objectives: The purpose of this study was to investigate the test-retest reliability, repeatability, and sensitivity of the modified shuttle test (MST) in adult patients with cystic fibrosis (CF).

Design: Prospective study.

Setting: Adult CF Unit, Belfast City Hospital.

Patients: Adult patients with CF.

Interventions: Test-retest reliability—none; sensitivity—inpatient IV antibiotic therapy for an acute exacerbation of respiratory disease.

Measurements: The test-retest reliability and repeatability of the MST was assessed by comparing performance on two consecutive MSTs performed in 12 patients with CF and stable disease. The sensitivity of the MST was assessed by measuring the change in MST performance after 2 weeks of IV antibiotic therapy in 24 patients admitted to hospital with acute exacerbations of their respiratory disease.

Results: In the assessment of test-retest reliability and repeatability (n = 12), there was a significant and strong correlation between trials for distance completed (Pearson’s r = 0.99; p < 0.01), peak heart rate (Pearson’s r = 0.99; p < 0.01), peak arterial oxygen saturation (SaO₂; Pearson’s r = 0.99; p < 0.01), and peak Borg rating of perceived breathlessness (Pearson’s r = 0.99; p < 0.01). The coefficients of repeatability for these variables were small (coefficient of repeatability: distance completed, 4 shuttles; peak heart rate, 6 beats/min; peak SaO₂, 4%; and peak Borg rating of perceived breathlessness, 0.9). In the assessment of sensitivity (n = 24), the standardized response mean (SRM) for distance completed on MST (SRM = 1.18) was the SRMs for spirometric measures of lung function (FEV₁, SRM = 0.96; FEV₁ percent predicted, SRM = 0.88).

Conclusions: This study demonstrates that the MST is a reliable, repeatable, and sensitive measure of exercise capacity in adult CF. The MST may be of value in determining prognosis, evaluation for lung transplantation, exercise prescription, and establishing the impact of new treatments on the disability associated with CF.

(CHEST 2000; 117:1666–1671)

Key words: cystic fibrosis; modified shuttle test; reliability; repeatability; sensitivity

Abbreviations: CF = cystic fibrosis; MST = modified shuttle test; SaO₂ = oxygen saturation; SRM = standardized response mean; SWT = shuttle walk test; V̇O₂peak = peak oxygen consumption

Exercise tests to evaluate exercise capacity are useful in determining prognosis, prescribing individualized exercise programs, and quantifying the disability associated with adult cystic fibrosis (CF) disease and its treatment.¹⁴ The ability to exercise in CF is related to the severity of lung disease, but there is great variability in the exercise capacity of patients with comparable lung function, and it is not possible to predict accurately a patient’s exercise tolerance from spirometric measures of lung function.²

A number of exercise tests have been used to directly assess exercise capacity in CF, although not all have been used in an adult population.⁵⁻⁸ Formal laboratory tests with on-line analysis of expired air provide detailed information about ability to exercise and, as for all populations, are considered the “gold standard” for exercise testing in CF. However, for-
ormal exercise tests are time-consuming and stressful, require specialized laboratory facilities and medical supervision, and are not widely available to many clinicians working in CF centers.

The alternative to formal exercise testing is informal exercise testing. The clinical utility of informal exercise tests in adult CF is determined by their ability to accurately measure exercise capacity at one particular point in time and their ability to respond to treatment-induced changes in exercise capacity. This can be established by analysis of the psychometric properties (validity, reliability, repeatability, and sensitivity) of the tests when used in an adult CF population.

We have developed a modified version of the shuttle walk test (SWT) for use in adults with CF.9,10 This test is a 15-level externally paced test and has been shown to overcome many of the problems associated with existing informal exercise tests.11 Performance on the modified shuttle test (MST) has been compared with peak oxygen consumption (V\text{O}_2\text{Peak}) measured during a treadmill test in a group of 20 adult patients with CF and varying degrees of lung function impairment.11 A strong relationship was found between V\text{O}_2\text{Peak} and MST with 90% of the variation of directly measured V\text{O}_2\text{Peak} explained by the variation in MST performance. These findings provide evidence of the criterion-related validity of the MST as a measure of exercise capacity in adult CF.

The aim of this study was to assess the test-retest reliability, repeatability, and responsiveness of the MST in CF and to investigate the relationship between spirometric measures of lung function and MST performance so that the clinical utility of the MST as a measurement tool of disability could be ascertained.

**Materials and Methods**

This study consisted of two separate experiments. In the first experiment, the test-retest reliability and repeatability of the MST was assessed by comparing performance on two consecutive MST tests in a group of adult patients with stable CF disease. In the second study, the responsiveness of the MST was determined by assessing the ability of the MST to detect changes in exercise capacity after a course of IV antibiotic treatment for an acute exacerbation of respiratory disease. The study was approved by the hospital ethical committee, and informed consent was obtained from all patients.

**Assessment of Test-Retest Reliability and Repeatability of the MST**

Twelve patients (9 men) were recruited for this experiment. All patients were clinically stable (<10% variation in FEV\textsubscript{1} from best during previous 6 months) and had been familiarized with both the lung function tests and the exercise test at the time of entry into the study. The patients made two visits to the hospital and were instructed not to administer bronchodilators for 3 h before each visit, but to take all other medications as usual. The mean interval between visits was 7 days (range, 4 to 14 days). Spirometric measures of lung function (Vitalograph Alpha; Vitalograph; Buckingham, United Kingdom), expressed as a percentage of predicted normal from the European Coal and Steel Workers Community Study,12 and an MST were performed at the same time of the day on each visit to the hospital.

**Assessment of Responsiveness of the MST**

Twenty-four patients (17 men) were recruited for this experiment. All patients were admitted to the hospital with an exacerbation of pulmonary disease. An acute exacerbation of pulmonary disease was defined as the presence of ≥4 of 12 signs and symptoms of an acute respiratory exacerbation (change in sputum production, new or increased hemoptysis, increased cough, increased dyspnea, malaise, fatigue or lethargy, fever, anorexia or weight loss, sinus pain or tenderness, changes in sinus discharge, drop in FEV\textsubscript{1} > 10% of previous value, and radiographic changes indicative of infection).13 Inpatient treatment consisted of parenteral antibiotic administration, intensified airway clearance, bronchodilator administration, and nutritional support. The choice of antibiotic depended on the specific organism isolated from sputum cultures. No patient was prescribed oral steroids during the course of this experiment. Spirometric measures of lung function (FEV\textsubscript{1} and FEV\textsubscript{1/2} percent predicted) and an MST were performed at approximately the same time of day at the beginning and at the end of IV antibiotic therapy. All patients were familiar with the lung function tests and the exercise test before entry into the study.

**The Modified Shuttle Test**

Baseline measurements included the Borg rating of perceived breathlessness,14 arterial oxygen saturation (SaO\textsubscript{2}), and resting heart rate (Ohmeda SaO\textsubscript{2} monitor with ear probe; Datex-Ohmeda; Madison, WI) and were recorded before each exercise test. The 15-level MST required the patients to walk/run at increasing speeds back and forth on a 10-m course. Patients were accompanied by an operator during the first minute of the test to help them pace themselves with the audio signal. At the end of each level, the patients were offered a standardized verbal encouragement (“good, keep going, you are doing well”).15 The patient was also told to go a little faster and reminded that they were permitted to run at any time during the test. Patients continued with the test until they were unable to do so or failed to maintain the set pace.16 Heart rate was measured at 15-s intervals using a short-range telemetry device (Polar Sports Tester; Polar Electro; Kempele, Finland). The end of the test was determined by the patient if they became unable to maintain the required speed or by the experimenter if the patient failed to complete a shuttle in the time allowed (that is, the patient was 0.5 m away from the cone when the beep sounded). At the end of the test, the peak heart rate, SaO\textsubscript{2}, and peak rating of perceived breathlessness were recorded.

**Statistical Analysis**

Descriptive statistics (mean ± SD) were calculated for all variables. The mean Borg scores were identical to median scores in this study, and exploratory data analysis showed that there were no outliers in the data. Mean ± SD of Borg ratings of perceived breathlessness have therefore been reported in preference to median scores for ease of interpretation.
Paired *t* tests were used to examine differences in heart rate, $\text{SaO}_2$, Borg rating of perceived breathlessness, and distance completed in the MST between study days. Pearson’s *r* and Spearman’s *p* correlation coefficients were used to assess the test-retest reliability of physiologic responses to exercise during the MST and to compare the relationship between spirometric measures of lung function and exercise capacity. Bland and Altman’s methods were used to calculate the limits of agreement between repeated MSTs (mean ± 2 SD for differences) and the coefficients of repeatability (twice the SD of differences). The standardized response mean (SRM) was used to assess responsiveness. The SRM is the ratio of change in average scores over time to the SD of change (mean scores at end of IV antibiotics [T2] − mean scores at beginning of IV antibiotics [T1]/SD of the change).17

RESULTS

Assessment of Test-Retest Reliability of the MST

Table 1 shows the descriptive statistics for the physical characteristics and baseline measurements on each study day. A diversity of patients in terms of age (range, 15 to 69 years) and FEV1 (range, 14 to 72% predicted) were included in this study, demonstrating that the patients exhibited a wide range of disease impairment. There was no significant difference in FEV1, resting $\text{SaO}_2$ and heart rate, or resting Borg rating of perceived breathlessness between the two study days (Table 1). This demonstrates that the patients were clinically stable throughout the study period.

There was a significant and strong correlation between trials for distance completed (Pearson’s *r* = 0.99; *p* < 0.01), peak heart rate (Pearson’s *r* = 0.99; *p* < 0.01), $\text{SaO}_2$ (Pearson’s *r* = 0.99; *p* < 0.01), and peak Borg rating of perceived breathlessness (Pearson’s *r* = 0.99; *p* < 0.01).

The mean shuttle distance completed, peak heart rates, and peak Borg rating of perceived breathlessness during the two trials are presented in Table 2. There was no significant difference in any of these variables between days 1 and 2. For each patient, the difference between days in MST performance is shown in Figure 1. The group mean difference (0 m) and the limits of agreement (−40 to +40 m) for the MST were small. The group mean difference and limits of agreement for the physiologic responses to exercise (Borg rating of perceived breathlessness, 0 [−1 to +1]; heart rate, −2 beats/min [−8 to +4 beats/min]; $\text{SaO}_2$, 1% [−3 to +5%]) were also small.

The coefficients of repeatability for the repeated MST were 40 m for change in distance completed, 6 beats/min for change in peak heart rate, 0.9 for absolute change in Borg rating of perceived breathlessness, and 4% for change in peak $\text{SaO}_2$.

Assessment of the Responsiveness of the MST

Table 3 summarizes the physical characteristics and the data for lung function and exercise performance at the beginning (T1) and at the end of IV antibiotics (T2). There was a significant improvement in both lung function and MST performance after IV antibiotic treatment.

The SRM for distance completed on MST was 1.18, for FEV1 was 0.96, and FEV1 percent predicted was 0.88. These values can be interpreted according to guidelines outlined by Meenan and colleagues. An SRM of approximately 0.20 is considered small, one of 0.50 indicates moderate responsiveness, and those of ≥0.80 are considered highly sensitive. As SRMs are standardized scores, the SRM for the MST can be compared directly with the SRMs for spirometric measures of lung function. These values show that the MST is highly responsive to IV antibiotic treatment-induced changes in dis-

---

**Table 1—Physical Characteristics and Baseline Measurements for MST on Day 1 and Day 2 (n = 12)**

<table>
<thead>
<tr>
<th>Baseline Measurement</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Mean Difference</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>(95% CI)</td>
<td></td>
</tr>
<tr>
<td>Age, yr</td>
<td>30.27 (15.33)</td>
<td>30.27 (15.33)</td>
<td>−0.02 (−0.05 to 0.01)</td>
<td>0.26</td>
</tr>
<tr>
<td>Height, m</td>
<td>1.70 (0.08)</td>
<td>1.70 (0.08)</td>
<td>−1 (−2 to 1)</td>
<td>0.7</td>
</tr>
<tr>
<td>FEV1, L</td>
<td>1.42 (0.80)</td>
<td>1.40 (0.79)</td>
<td>0 (−1 to 1)</td>
<td>0.88</td>
</tr>
<tr>
<td>FEV1, % predicted</td>
<td>40 (20)</td>
<td>40 (19)</td>
<td>0 (−1 to 1)</td>
<td>0.34</td>
</tr>
<tr>
<td>Resting heart rate, beats/min</td>
<td>95 (9)</td>
<td>94 (8)</td>
<td>−1 (−2 to 1)</td>
<td>0.7</td>
</tr>
<tr>
<td>Resting $\text{SaO}_2$, %</td>
<td>92 (3)</td>
<td>92 (2)</td>
<td>0 (−1 to 1)</td>
<td>0.88</td>
</tr>
<tr>
<td>Resting Borg score</td>
<td>0 (1)</td>
<td>0 (1)</td>
<td>0.08 (−0.1 to 0.3)</td>
<td>0.58</td>
</tr>
</tbody>
</table>

*Results are reported as mean ± SD.
ability and that its sensitivity is better than spirometric measures of lung function.

Relationship Between Exercise Performance and Spirometric Measures of Lung Function

The relationship between FEV₁ percent predicted and MST distance was moderate in both experiment 1 and 2, although there was a large variation in MST performance in patients with comparable lung function. There was a poor correlation between change in FEV₁ and change in MST (Table 4).

Discussion

This study shows that the MST is a highly reproducible, repeatable, and responsive measure of exercise capacity in an adult CF population. The reproducibility of the SWT has been previously assessed in COPD and in cardiac populations.9,14 The original authors of the test found that a practice test was necessary to ensure reliable results in COPD patients.9 This finding has important implications regarding the clinical utility of the SWT as a routine measure of exercise capacity because clinicians can only be confident that their results are reliable if they allow patients to perform at least one practice test. This means that each exercise test could take up to 45 min to complete (two SWTs + preparation time + rest period between tests), time not available to most clinicians working in adult CF centers. Furthermore, it is unlikely that adult patients with CF would be motivated enough to comply with protocols that involve practice exercise tests.

In contrast to the original authors’ work, Arnott et al14 found that there was no significant improvement in exercise performance when the SWT was repeated in a cardiac population and suggested that practice SWTs were not necessary to obtain a reliable measure of exercise capacity in cardiac populations.

The results of this present study also suggest that routine practice tests are not necessary when using the MST to evaluate exercise capacity in adult patients with CF who have been previously familiarized with the test. There was a highly significant correlation in distance walked and in the physiologic response to exercise between tests, suggesting that practice tests are not necessary to obtain a reliable measure of exercise capacity. The corresponding coefficients of repeatability were also small, suggesting that the MST is a repeatable measure of exercise capacity in adult CF. These results are important for two reasons. First, the findings suggest that exercise capacity may potentially be predicted from a single MST with accuracy. Second, the results permit us to suggest the minimum amount of change on the MST that indicates an important change in clinical condition. Performance on the MST must improve by > 4 shuttles (40 m) to ensure that the change in performance is not related to measurement error in the instrument. Therefore, if a treatment-induced change in MST is > 40 m, it is probable that the intervention has had an effect on exercise perfor-

Table 2—Distance Completed and Physiologic Responses to Exercise for MST on Day 1 and Day 2 (n = 12)*

<table>
<thead>
<tr>
<th>Measure of Exercise Performance</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Mean Difference (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance completed, m</td>
<td>754 (361)</td>
<td>754 (362)</td>
<td>0 (–1 to 1)</td>
<td>0.98</td>
</tr>
<tr>
<td>Peak heart rate, beats/min</td>
<td>162 (32)</td>
<td>160 (33)</td>
<td>–2 (–3 to 1)</td>
<td>0.11</td>
</tr>
<tr>
<td>Peak Borg score</td>
<td>6 (2)</td>
<td>6 (2)</td>
<td>0 (–1 to 1)</td>
<td>0.12</td>
</tr>
<tr>
<td>SaO₂, %</td>
<td>89 (4)</td>
<td>90 (5)</td>
<td>1 (–1 to 2)</td>
<td>0.16</td>
</tr>
</tbody>
</table>

*Results are reported as mean ± SD.
mance. Analysis of the difference between trials in distance covered, peak heart rate, and peak Borg rating of perceived breathlessness provide further evidence that practice tests are not necessary to obtain a repeatable measure of exercise capacity in CF.

This study has also shown that the MST was highly sensitive to IV antibiotic changes in disability and that the sensitivity of the MST was greater than routine spirometric measures of lung function. This has important clinical implications for evaluating the outcome of new treatments, particularly in adult patients with more severe disease impairment. In patients with severe disease impairment (FEV1 < 25% predicted), spirometry is less useful than other measures of clinical status in diagnosing an acute exacerbation of respiratory disease or in assessing the outcome of an intervention. The MST may provide a much needed objective and sensitive measure of disease status or the outcome of an intervention in such patients.

Although there was a moderate relationship between spirometric measures of lung function (FEV1 percent predicted) and MST performance in both experiments in this study, a large variability in exercise performance in patients with comparable lung function still existed. Furthermore, treatment-induced changes in MST performance related poorly to treatment-induced changes in lung function.

These findings confirm the results of previous studies and provide further confirmation that the disability associated with CF and its treatment cannot be predicted with any accuracy from lung function measurements alone and that direct measurement of exercise capacity is necessary to objectively quantify the extent of a patient’s disability.2

We have provided evidence of the validity, reliability, repeatability, and sensitivity of the MST as a measure of peak exercise capacity in adult patients with CF. The MST may be of value in determining prognosis, evaluation for lung transplantation, exercise prescription, and establishing the impact of new treatments on the disability associated with CF.

Table 3—Physical Characteristics, Spirometric Measures of Lung Function, and Exercise Performance at T1 and T2 (n = 24)*

<table>
<thead>
<tr>
<th>Measurement</th>
<th>T1 Mean (SD)</th>
<th>T2 Mean (SD)</th>
<th>Mean Difference (95% CI)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, yr</td>
<td>31 (10)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Height, m</td>
<td>1.72 (0.10)</td>
<td>1.83 (1.04)</td>
<td>0.30 (0.17 to 0.44)</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>FEV1, L</td>
<td>1.53 (0.82)</td>
<td>1.83 (1.04)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>FEV1, % predicted</td>
<td>42 (20)</td>
<td>50 (26)</td>
<td>8 (4 to 11)</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Distance completed</td>
<td>692 (289)</td>
<td>867 (336)</td>
<td>175 (112 to 237)</td>
<td>&lt; 0.01</td>
</tr>
</tbody>
</table>

*Results are reported as mean ± SD.

Table 4—Correlation Between FEV1 Percent Predicted and Distance Completed on MST

<table>
<thead>
<tr>
<th>Study</th>
<th>Spearman’s ρ</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 1</td>
<td>0.70</td>
<td>0.02</td>
</tr>
<tr>
<td>Day 2</td>
<td>0.65</td>
<td>0.05</td>
</tr>
<tr>
<td>Sensitivity study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>0.68</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>T2</td>
<td>0.67</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>T2 - T1</td>
<td>0.05</td>
<td>0.81</td>
</tr>
</tbody>
</table>

References

10. Singh SJ, Morgan MDL, Hardman AE, et al. Comparison of oxygen uptake during a conventional treadmill test and the...
17 Garrat AM, Ruta DA, Abdalla MI, et al. SF36 Health Survey Questionnaire: II. Responsiveness to changes in health status in four common clinical conditions. Qual Health Care 1994; 3:186–192

Clinical Pulmonary Medicine

Don't miss these sessions offered from the Clinical Pulmonary Medicine track at Chest 2000.

- COPD: Global Initiative (GOLD)
- Emphysema—Not the Same Old Story
- Hemoptysis: Case-Based Approach
- Angiogenesis of the Respiratory Muscles: Animal Experiment with Clinical Application
- Diagnosis and Treatment of Pulmonary Hypertension
- Sinus Disease/Sinus Infections Related to Pulmonary Medicine
- Standardizing Bronchial Provocation Testing
- Heparin and Heparinoids
- Inhaled Corticosteroids Should Be Used in COPD
- Nonpulmonary Emergencies
- Noninvasive Positive Pressure Home Ventilation for COPD
- Smoking Cessation
- Controversial Issues With Inhaled Corticosteroids

For the complete list of sessions call: 800-343-ACCP or visit: www.chestnet/CHEST/2000/