A Step in the Right Direction: Assessing Exercise Tolerance in Cystic Fibrosis

Ian M. Balfour-Lynn, MD, MRCP, FRCS(Ed), FRCPCH,1* S. Ammani Prasad, MCSP,1 Aidan Laverty, BSC,1 Bruce F. Whitehead, FRCP, FRCPCH,2 and Robert Dinwiddie, FRCP, FRCPCH1

Summary. Exercise tolerance may be reduced in patients with cystic fibrosis, but it is not always possible to predict this from standard lung function measurements. Formal exercise testing may, therefore, be necessary, and the test should be simple and readily available. We have developed a "3-minute step test" and compared it with the standard 6-minute walking test. Subjects stepped up and down a 15-cm-high single step at a rate of 30 steps per minute for 3 minutes. The effect of the step test on spirometry was tested first in 31 children with CF (mean age, 12.0 years), who had a mean (range) baseline forced expired volume in 1 second (FEV1) of 64% (18–94%) of predicted values. The step test was then compared with the standard 6-minute walk in a further 54 patients with cystic fibrosis (mean age, 12.5 years), with mean (range) baseline FEV1 of 61% (14–103%) of predicted values. Outcome measures were minimum arterial oxygen saturation (SaO2), maximum pulse rate, and the modified Borg dyspnea score.

Post-step test spirometry showed mean (95% CI) changes of −1.1% (−6.0+3.9%) for forced vital capacity, of −1.6% (−4.2+1.1%) for FEV1, and +0.25% (−2.8+3.3%) for peak expiratory flow, although 5/31 children showed >15% drop in one or more parameters. The step and walk tests both produced significant changes (P < 0.0001) in all outcomes, with a mean (range) minimum SaO2 of 92% (75–98%) versus 92% (75–97%), a maximum pulse rate of 145 b.p.m. (116–189) versus 132 (100–161), and a Borg score of 2.5 (0–9) versus 1.0 (0–5), respectively. Comparison of the two tests showed that the step test increased breathlessness (mean change Borg score, 2.3 vs. 0.8; P < 0.0001) and pulse rate (mean change, 38% vs. 24%, P < 0.0001) significantly more than the walk, whereas the decrease in SaO2 was similar (mean change, −2.9% vs. −2.6%; P = 0.12). Some patients with a significant drop in SaO2 (>4%) would not have the decrease predicted from their baseline lung function. Reproducibility for the two tests was similar. The step test is quick, simple and portable, and is not dependent on patient motivation. Although the step test is more tiring, its effect on SaO2 is similar to the 6-minute walking test. It is a safe test that may prove to be a valuable measure of exercise tolerance in children with pulmonary disease, although longitudinal studies are now needed. Pediatr Pulmonol. 1998; 25:278–284.

© 1998 Wiley-Liss, Inc.

Key words: exercise testing; exercise tolerance; cystic fibrosis.

INTRODUCTION

With the inevitable progression of lung disease in cystic fibrosis (CF), the patient’s ability to exercise becomes reduced.1 Although this reduction is associated with declining lung function, it is not always possible to predict exercise tolerance from standard lung function measurements.2,3 Assessment of fitness and exercise tolerance in cystic fibrosis is a useful measure of the impact the disease is having on the patient, particularly when repeated over time.4 Currently there are several options for exercise testing, but they all have problems. We have, therefore, developed the three-minute step test, modified from the original Master two-step exercise test used for adult cardiac assessment (reviewed by Naughton5). Subjects step up and down on a single level platform for a fixed time at a fixed rate. The advantages of the test are that it is quick, simple, and portable and is not dependent

1Respiratory Unit, Great Ormond Street Hospital for Children, London, United Kingdom.
2Cardio-Thoracic Unit, Great Ormond Street Hospital for Children, London, United Kingdom.
*Correspondence to: Dr. I.M. Balfour-Lynn, Paediatric Respiratory Medicine, Royal Brompton Hospital, Sydney Street, London SW3 6NP, UK.

Received 6 June 1997; accepted 11 December 1997.

© 1998 Wiley-Liss, Inc.
on patient motivation. We have evaluated the step test and compared it with the standard 6-minute walk for assessment of symptomatic exercise tolerance in children with cystic fibrosis.

MATERIALS AND METHODS

The study consisted of three parts: (1) the effect of the step test on lung function, carried out as a safety measure to check that large falls in lung function were not produced during the test (part 1); (2) comparison of the step test with the standard 6-minute walking test (part 2); and (3) assessment of reproducibility of both tests (part 3). The study was approved by the Hospital Research Ethics Committee, and parents and/or patients gave written informed consent.

Subjects

Subjects were children with CF attending an outpatient clinic or undergoing heart-lung transplantation assessment. Exclusion criteria were age under 6 years, concomitant diagnosis of asthma, current oxygen therapy, or a resting oxygen saturation ($\text{SaO}_2$) less than 85%.

Part 1

There were 31 patients (15 girls and 16 boys) with a mean age of 12.0 years (range, 6–17 years). Their mean baseline forced expiratory volume in 1 second ($\text{FEV}_1$) was 64% of predicted for gender and height (range, 18–94%), and their mean resting $\text{SaO}_2$ was 95% (range, 89–98%).

Part 2

There were 54 patients (32 girls and 22 boys) with a mean age of 12.5 years (range, 6–18 years). Their mean baseline $\text{FEV}_1$ was 61% of predicted (range, 14–103%), and their mean resting $\text{SaO}_2$ was 95% (range, 87–98%).

Part 3

Twelve of the patients (8 girls, 4 boys, with mean age 13.6 years) from part 2 also took part in the reproducibility study. Subjects who did repeat tests were in-patients at the end of a stay for intravenous antibiotic therapy. Their lung function was no different from the group as a whole (mean baseline $\text{FEV}_1$ 64% of predicted and mean resting $\text{SaO}_2$ 95%), and they were clinically stable.

Study Design

Part 1

The subjects performed the step test only. Lung function was assessed before and after exercise by standard spirometry using a compact spirometer (Vitalograph, Buckingham, UK). Forced vital capacity (FVC) and $\text{FEV}_1$ were recorded as percentage predicted for gender and height. Peak expiratory flow (PEF) was measured with a Wright’s peak flow meter. A maximum of three measurements was recorded for each parameter. Resting $\text{SaO}_2$ was measured using a Biox 3700e pulse oximeter (Ohmeda, Boulder, CO USA) with a flexiprobe attached to the finger and its lead strapped to the subject’s arm.

Part 2

Subjects were randomized for order of testing. Baseline spirometry and $\text{SaO}_2$ were measured; the patients then trained, practiced both tests, and then rested. The first test was performed; the second one was started 30 minutes after pulse rate and $\text{SaO}_2$ had returned to baseline pre-exercise levels. Standardized encouragement was offered during the tests. All tests were supervised by the same investigator. Outcome measures were minimum $\text{SaO}_2$ and maximum pulse rate during exercise, and the modified Borg breathlessness score at the end of exercise (see Appendix). $\text{SaO}_2$ and pulse rate were measured by pulse oximetry as described above. Traces were recorded on a computer program for later review, which helped eliminate false readings due to low-quality signals or movement artefacts.

The step test. Subjects stepped up and down a commercially available single step set at a height of 15 cm (6 inches). The stepping rate of 30 per minute for 3 minutes was controlled by a metronome. Patients could stop if they felt tired or if the $\text{SaO}_2$ fell below 75%, in which case the total number of steps taken was calculated. Patients were both encouraged and shown how to change the leading leg to reduce localized muscle fatigue.

The 6-minute walking test. Subjects walked up and down a measured 17-m corridor for 6 minutes, accompanied by the investigator with the pulse oximeter on a trolley. The test is self-paced and subjects could rest when they wished but the clock continued to run. The test was abandoned if the $\text{SaO}_2$ fell below 75%. Total distance walked was recorded.

Part 3

Some subjects performed both tests on each of 2 consecutive days. The tests were performed at approximately the same time of day and in the same order.

Abbreviations

- bpm: Beats per minute
- CF: Cystic fibrosis
- CI: Confidence interval
- $\text{FEV}_1$: Forced expiratory volume in 1 second
- FVC: Forced vital capacity
- $\text{PaO}_2$: Partial pressure of oxygen in arterial blood
- PEF: Peak expiratory flow
- $\text{SaO}_2$: Arterial oxygen saturation
- SD: Standard deviation
Statistical Analysis

Statistical analysis was performed using Minitab software (Minitab Inc., State College, PA). Paired t tests were used to compare the outcomes before and after exercise. Values of $P < 0.05$ were considered significant. Bland-Altman plots (difference between the methods versus mean for the two methods) were used to compare the two tests (step test and 6-minute walk). Reproducibility of the tests was assessed using Bland-Altman plots to compare results for the repeat tests. The coefficient of repeatability was calculated using the British Standards Institution definition, i.e., twice the standard deviation of the differences.

RESULTS

Part 1: Effect of the Step Test on Lung Function

For the group, there was no significant change in lung function (Table 1), with as many children showing an improvement as showing a decline after exercise. Lung function remained within ±15% in 26 of the children and decreased by more than 15% in 5 children. The effect of exercise could not have been predicted in these five children from their baseline FEV$_1$ (18, 24, 43, 50, and 58% predicted), although they were all below the mean baseline for the group.

Part 2: Effect on $\text{SaO}_2$, Pulse Rate, and Breathlessness

The step test

There was a significant fall in $\text{SaO}_2$, rise in pulse rate and rise in Borg score ($P < 0.0001$) during exercise (Tables 2 and 3; Fig. 1). All subjects completed the walk without stopping for a rest, with a mean walking distance of 422 m (range, 272–560 m). There was a degree of correlation between walking distance and baseline FEV$_1$ ($r = 0.49$; $P < 0.0001$) or resting $\text{SaO}_2$ ($r = 0.35$; $P < 0.001$).

Comparison of the two tests

There were no significant differences in baseline $\text{SaO}_2$, pulse rate, and Borg score for the subjects before the start of each of their two tests. Mean (range) resting $\text{SaO}_2$ was 95% (87–98%), pulse rate was 106 (71–139), and Borg score 0.2 (0–2). The step test produced a significantly greater increase in pulse rate and Borg score than the walk ($P < 0.0001$; Table 4). Although the fall in $\text{SaO}_2$ was comparable between the two tests (Table 4; Fig. 3), the limits of agreement for the change in $\text{SaO}_2$ were i.e., 90 steps. The five who did not, managed 45, 45, 50, 52, and 60 steps. Reasons for stopping were low $\text{SaO}_2$ (<75%), excessive breathlessness, or tired legs. There were 13 patients with a fall in $\text{SaO}_2$ of 4% or more during exercise (five during step test only, three during walk only, and five during both tests).

Although as a group the patients had a significantly decreased baseline FEV$_1$ and $\text{SaO}_2$, it was not possible to predict which individuals would desaturate the most, because some had a reasonable baseline FEV$_1$, and many had a normal resting $\text{SaO}_2$. However, the lower the baseline $\text{SaO}_2$, the greater the fall during exercise tended to be (Fig. 2).

The 6-minute walk

There was a significant fall in $\text{SaO}_2$, rise in pulse rate and rise in Borg score ($P < 0.0001$) during exercise (Tables 2 and 3; Fig. 1). All subjects completed the walk without stopping for a rest, with a mean walking distance of 422 m (range, 272–560 m). There was a degree of correlation between walking distance and baseline FEV$_1$ ($r = 0.49$; $P < 0.0001$) or resting $\text{SaO}_2$ ($r = 0.35$; $P < 0.001$).

Comparison of the two tests

There were no significant differences in baseline $\text{SaO}_2$, pulse rate, and Borg score for the subjects before the start of each of their two tests. Mean (range) resting $\text{SaO}_2$ was 95% (87–98%), pulse rate was 106 (71–139), and Borg score 0.2 (0–2). The step test produced a significantly greater increase in pulse rate and Borg score than the walk ($P < 0.0001$; Table 4). Although the fall in $\text{SaO}_2$ was comparable between the two tests (Table 4; Fig. 3), the limits of agreement for the change in $\text{SaO}_2$ were

| TABLE 1—Percentage Change in Percent Predicted FVC, FEV$_1$, and PEF in 31 Subjects After the Step Test |
|----------|--------|--------|
| FVC (%)  | −1.1 ± 0.6 to +3.9 |
| FEV$_1$ (%) | −1.6 ± 1.0 to +1.1 |
| PEF (%)  | +0.25 ± 0.28 to +3.3 |

| TABLE 2—Comparison of Minimum Oxygen Saturation ($\text{SaO}_2$, %), Maximum Pulse Rate (bpm), and Maximum Borg Score at Rest and During Step Test and 6-Minute Walk, in 54 Subjects$^1$ |
|----------|--------|--------|--------|
| Resting | Step test | Walk |
| Minimum $\text{SaO}_2$ (%) | 95 (94–96) | 92 (91–94) | 92 (91–94) |
| Maximum pulse rate (bpm) | 106 (103–110) | 145 (141–148) | 132 (128–135) |
| Maximum Borg score | 0.2 (0.1–0.3) | 2.5 (2.0–3.1) | 1.0 (0.7–1.3) |

$^1$Results are shown as means (95% CI).
relatively wide (−3.2+4.0%), so the two tests are not interchangeable. The relationship between maximum fall in \(SaO_2\) and baseline FEV\(_1\) was similar for the step test and walk (\(r = 0.52\) and \(r = 0.51\), respectively). Correlation of maximum fall in \(SaO_2\) with resting \(SaO_2\) was slightly better for the step test than the walk (\(r = 0.74\) and \(r = 0.65\), respectively).

Part 3: Reproducibility of Step Test and 6-Minute Walk

The reproducibility of duplicate step tests (n = 12) and 6-minute walks (n = 9) in terms of change in \(SaO_2\), pulse rate, and Borg score after exercise were found to be similar and clinically acceptable (Table 5). This should, however, be put into the context of the small numbers used for this part of the study. The coefficients of repeatability (as assessed by the Bland-Altman method) for the duplicate step tests and 6-minute walks respectively were comparable, namely 2.3 versus 1.5 for percentage change in \(SaO_2\), 34 versus 36 for percentage change in pulse rate, and 1.4 versus 1.5 for absolute change in the modified Borg score. Mean difference for distance walked during the repeat tests (first minus second) was 8.3 m (limits of agreement, −35 to +18 m). Number of steps taken was identical in 11 of 12 subjects, with a difference of five steps in the remaining subject.

DISCUSSION

The 3-minute step test was harder exercise than the 6-minute walking test, as shown by the significantly higher pulse rate and breathlessness score obtained. It was expected that the harder exercise would produce a greater effect on \(SaO_2\), but this was not the case, although the drop in \(SaO_2\) did occur over a shorter time period with the step test. The reason for this may have been that the harder exercise produced a greater degree of anaero-
patients tend to find their own pace, and their maximal exercise test in the true sense. In the walk, there were fewer of these during the step test than the walk. This difference was because the subjects’ arms could be kept relatively still during the step test, whereas their movement artifacts in some recordings, although more recently use of pulse oximetry in exercise testing. Different oximeters have varying degrees of accuracy under exercise conditions as shown by Gaskin and Thomas. Location of the probe is another factor that affects the use of pulse oximetry in exercise testing. Different oximeters have varying degrees of accuracy under exercise conditions. As is common practice in exercise testing, pulse oximetry was used to measure \( \text{SaO}_2 \) because it is simple and noninvasive. The \( \text{SaO}_2 \) is measured as a substitute for the \( \text{PaO}_2 \), and their relationship is not linear due to the shape of the oxyhemoglobin dissociation curve. For a given fall in \( \text{PaO}_2 \), the measured fall in \( \text{SaO}_2 \) is greater when starting at a lower baseline \( \text{PaO}_2 \). For example, a fall in \( \text{PaO}_2 \) from 12 to 11 kPa (90 to 83 mmHg) leads to a fall in \( \text{SaO}_2 \) of 1.5%, whereas a fall from 8 to 7 kPa (60 to 53 mmHg) leads to a fall in \( \text{SaO}_2 \) of 4%. This finding means that the effect of exercise on \( \text{SaO}_2 \) is greater for those with worse lung disease, and even less oxygen is made available to their tissues. In addition, patients with the most damaged lungs experience a greater drop in \( \text{PaO}_2 \) during exercise due to impaired oxygen diffusion. In this study, the drop in \( \text{SaO}_2 \) was dramatic in some patients (Fig. 2), and they tended to be the ones with such severe lung disease that they were undergoing assessment for heart-lung transplantation.

Whereas there is a need for caution when interpreting changes in \( \text{SaO}_2 \), other problems are associated with the use of pulse oximetry in exercise testing. Different oximeters have varying degrees of accuracy under exercise conditions as shown by Gaskin and Thomas. In general, oximeters are more accurate in the higher saturation ranges and tend to overestimate in the lower ranges (\( \text{SaO}_2 < 90\% \)). This means that they may be reliable when used in healthy subjects, but drops in \( \text{SaO}_2 \) may be underestimated in patients with cardiorespiratory problems. Measurements of pulse rates are also limited in the higher ranges, and an underestimation may be seen with heavy exercise.

Severe finger clubbing does not seem to cause problems with finger probes. To minimize problems with signals in this study, the flexiprobe was taped to the index finger, and the lead was strapped to the arm to reduce movement artifact. False readings were also eliminated after each test because the signals were continuously displayed and recorded alongside the readings. Despite this, there were still movement artifacts in some recordings, although there were fewer of these during the step test than the walk. This difference was because the subjects’ arms could be kept relatively still during the step test, whereas the arms tended to swing when walking. In addition, the oximeter and its lead were kept stationary during the step test rather than being wheeled up and down the corridor.

Exercise testing has an important role in assessing patients with CF. In the United Kingdom, current standards
for CF care now recommend regular annual exercise testing for all patients.\textsuperscript{15} Spirometry and $SaO_2$ performed at rest are of course useful, but are not always good predictors of how well patients can exert themselves as part of normal daily life.\textsuperscript{2,3} The 12-minute walk was originally described to assess elderly patients with chronic obstructive pulmonary disease because this was felt to mirror everyday activities.\textsuperscript{16,17} It is often useful to assess exercise tolerance in children to gain an insight into how well they can walk up stairs at school or even run for a bus. It can also reassure parents and teachers that a child is fit enough to play games and sports, and the child can then be encouraged to adopt an active lifestyle.\textsuperscript{4} Some patients in this study who had reasonable lung function and $SaO_2$ at rest, had significant oxygen desaturation during exercise, which would not have been predicted. Exercise-induced desaturation is not limited to patients with advanced lung disease, as has been previously suggested.\textsuperscript{1} Aerobic fitness, as assessed by cycle ergometry, has also been shown to be a valid long-term prognostic indicator in CF.\textsuperscript{18} It is too soon to say whether simpler exercise tests will also prove to be useful for this purpose in CF, although the walking test has been used to predict long-term outcome in cardiac disease.\textsuperscript{19} In some centers, the walking test is currently used as part of the prognostic assessment for suitability for heart-lung transplant in children with CF.\textsuperscript{20} A simple test of exercise tolerance would also make a useful outcome measure for therapeutic trials and can help assess the effects of many forms of intervention.

To be really useful as a clinical tool, an exercise test must be simple to perform and require minimal training for both the patient and the supervisor. It should be one that can be used easily in a clinic, on a ward, or in a respiratory laboratory. It should also be portable so that it can be used in field work, for example when studying normal children in their homes or at school. Currently there are several options for exercise testing in children. Tests of maximal exercise tolerance on a cycle ergometer or treadmill with continuous respiratory gas analysis give the most information and are considered as the gold standard. Recent evidence suggests that this form of exercise may not be appropriate because it is not representative of physical activity normally undertaken by children.\textsuperscript{21} In addition, these tests are time-consuming, stressful, not available outside a specialist laboratory and require medical supervision. Twelve- or six-minute walking tests have been well validated in adults,\textsuperscript{16,22,23} but although widely used, have only recently been assessed in children.\textsuperscript{24,25} Tests of shorter duration are said to be equally valid\textsuperscript{22} and significant improvements in the 2-minute walking distance were seen in children with CF treated for chest exacerbations.\textsuperscript{26} However, these tests require adequate space and are awkward; they are also effort-dependent so the patient’s attitude and motivation are a major factor in determining the distance walked.\textsuperscript{27,28} For this reason our clinical practice is to use the fall in $SaO_2$ as the critical outcome measure, rather than the distance walked during the 6 minutes. This strategy means that although the step test has no equivalent discriminator of work performed for most subjects (because the majority complete 3 minutes and hence perform an equal number of steps), it does not put the step test at a disadvantage. Finally there is the shuttle walking test, which is a standardized incremental test; it requires adequate space and training and has not yet been validated in children.\textsuperscript{29,30}

Various forms of step test have been described\textsuperscript{31–37} since the original one devised by Master and Oppenheimer.\textsuperscript{38} Only one form has been validated in children\textsuperscript{39} and most are designed for assessing fitness in healthy adults. Differences between protocols are mainly related to step height and frequency, and whether these parameters are kept constant or graded. In many, the step was set at levels too high for younger children. We found that the step height of 15 cm produced an adequate level of exercise for patients with cardiorespiratory compromise. The stepping height and frequency in our study were identical for all subjects and kept the same throughout the test. The work performed in the test is a product of the stepping height and rate as well as the subjects weight and height (particularly leg length).\textsuperscript{32,36} To keep the amount of work constant between subjects, the height or rate could be adjusted to account for differences in patient size. In practice this adjustment is neither feasible, nor important in the context of the way the test is used, although the pubertal growth spurt might affect interpretation of longitudinal results from annual testing through puberty. Some of the children who were not used to any exercise found the step test caused localized leg muscle fatigue, which was less of a problem with the walk. For any patients with joint problems, a soft mat on the floor adjacent to the step is recommended, although clearly the step test should be avoided in those with severe CF-related arthropathy affecting their lower limbs.

In conclusion, this study has evaluated the step test as a means of assessing exercise tolerance in children. To differentiate our methods from those of other step tests, it is suggested that our exercise test be known as the “3-minute step test.” It is quick, simple, portable, and requires little space. Importantly, the step test also excludes patient motivation as a factor in determining outcome. It is a safe test that may prove to be valuable for both clinical and research purposes, but longitudinal studies are necessary to assess the effects of time and therapeutic interventions on step test performance.

**ACKNOWLEDGMENTS**

We thank all the children who took part in this study. We also thank Dr. Janet Stocks and Dr. Rod Lane for their helpful comments.
APPENDIX

Modified Borg Dyspnea Score:
0, nothing at all; 0.5, very, very slight (just noticeable); 1, very slight; 2, slight; 3, moderate; 4, somewhat severe; 5, severe; 7, very severe; 9, very, very severe (almost maximal); 10, maximal. Numbers 6 and 8 have no accompanying descriptive term.

REFERENCES